THE ANNALS *of* MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. XIV, No. 2 — June, 1943

THE ANNALS OF MATHEMATICAL STATISTICS

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves., Baltimore, Md., or to the Secretary of the Institute of Mathematical Statistics, E. G. Olds, Carnegie Institute of Technology, Pittsburgh, Pa. Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue. The months of issue are March, June, September and December. Because of war-time difficulties of publication, issues may often be from two to four weeks late in appearing. *Subscribers are therefore requested to wait at least 30 days after month of issue before making inquiries concerning non-delivery.*

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $5.00 per year. Single copies $1.50. Back numbers are available at $5.00 per volume, or $1.50 per single issue.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879

ON TRANSFORMATIONS USED IN THE ANALYSIS OF VARIANCE

By J. H. Curtiss

Cornell University

1. Introduction. Transformations of variates to render their distributions more tractable in various ways have long been used in statistics [12, chapter XVI]. The present extensive use of the analysis of variance, particularly as applied to data derived from designs such as randomized blocks and Latin squares, has placed new emphasis on the usefulness of such transformations. In the more usual significance tests associated with the analysis of variance, it is assumed *a priori* that the plot yields are statistically independent normally distributed variates which all have the same variance, but which have possibly different means. The hypotheses to be tested are then concerned with relations among these means. But in practice, it sometimes seems appropriate to specify for each variate a distribution in which the variance depends functionally upon the mean; moreover, in such cases, the specification is generally not normal. For example, when the data is in the form of a series of counts or percentages, a Poisson exponential or binomial specification may seem in order, and the variance of either of these distributions is functionally related to the mean of the distribution. Before applying the usual normal theory to such data, it is clearly desirable to transform each variate so that normality and a stable variance are achieved as nearly as possible.

Various transformations have been devised to do this, and a number of articles explaining the nature and use of these transformations have recently been published.[1] However, the available literature on the subject appears to be mainly descriptive and non-mathematical. The object of this paper is to provide a general mathematical theory (sections 2 and 3) for certain types of transformations now in use. In the framework of this theory we shall discuss in particular the square root and inverse sine transformations (section 4), and also several logarithmic transformations (section 4 and section 5).

2. General theory. As it arises in the analysis of variance, the problem of stabilizing a variance functionally related to a mean may be stated as follows: Suppose X is a variate whose mean $\mu = E(X)$ is a real variable with a range S of possible values, and whose standard deviation $\sigma = \sigma_X = \sigma(\mu)$ is a function of μ not identically constant. Required, to find a function $T = f(X)$ such that both $f(X)$ and $\sigma_T^2 = E\{[T - E(T)]^2\}$ are functionally independent of μ for μ on S. (By "functionally independent," we mean that $\frac{\partial f}{\partial \mu} \equiv 0$, and $\frac{\partial \sigma_T^2}{\partial \mu} \equiv 0$ for μ on S.)

[1] See references [1], [2], [3], [4], [5], [6], [13], [16].

The following line of argument is adopted in certain of the references mentioned above ([1], [2], [3], [4]): From the relation $dT = f'(X)dX$, we deduce as an approximation by some sort of summation process that $\sigma_T = f'(\mu)\sigma(\mu)$. Setting this expression equal to a constant, say c, we obtain $f'(\mu) = c/\sigma(\mu)$, so $f(x)$ is an indefinite integral of $c/\sigma(x)$. The roughness of the approximation used here is only too apparent.[2] For example, if X is normally distributed, then the variance of $T = X^2$ as given by the approximation is $4\sigma^2\mu^2$, while actually it is $4\sigma^2\mu^2 + 2\sigma^4$.

Indeed, it is easily seen that in important special cases the problem of stabilization as above stated could have no solution other than the trivial one in which T is identically constant on the set of points of increase of the d.f.[3] of X. For instance, if X has a Poisson exponential distribution, then the identity $E[\{f(X) - E[f(X)]\}^2] \underset{\mu}{\equiv} c$, or $E\{[f(X)]^2\} \underset{\mu}{\equiv} c + \{E[f(X)]\}^2$, becomes

$$\sum_{k=0}^{\infty} [f(k)]^2 \frac{e^{-\mu}\mu^k}{k!} \underset{\mu}{\equiv} c + \left[\sum_{k=0}^{\infty} [f(k)] \frac{e^{-\mu}\mu^k}{k!}\right]^2, \qquad \mu > 0.$$

Expanding both sides in powers of μ, we need only equate the coefficients of the zero-th power of μ on each side to find that $[f(0)]^2 = c + [f(0)]^2$, which implies that $c = 0$ and hence that $f(0) = f(1) = f(2) = \cdots$. A similar demonstration can be given for the case in which X has a binomial distribution with a fixed number of values of the variate.

As to the problem of choosing $T = f(X)$ so that its distribution is exactly normal, we can observe immediately that a single-valued function $f(X)$ will never transform a variate X with a discrete distribution into a variate with a continuous one. On the other hand, any variate X with a continuous d.f. $F(x)$ can be transformed into a normally distributed variate T by the transformation $T = f(X)$ defined by the equation

$$F(X) = \int_{-\infty}^{T} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\, dt.$$

However, aside from the practical difficulty of solving this equation for T, the resulting function $T = f(X)$ will not generally be functionally independent of the mean of X.

These considerations lead us to seek asymptotic solutions to the problems of normalization and stabilization. Such solutions are considered in the next section.

3. Asymptotic theorems. In the remainder of this paper, we shall suppose that the distribution of X depends on a parameter n which is to tend somehow to

[2] Tippett [14] says: "This derivation is not mathematically sound, and the result is only justified if on application it is found to be satisfactory."

[3] i.e., distribution function. For any given one-dimensional variate X we shall denote the probability or relative frequency assigned to a set R by $P(R)$. The d.f. of the variate then is the point function $F(x) = P(X \leqq x)$. This function is sometimes called the cumulative frequency function of X.

infinity. The mean $\mu = \mu_n$ of X, with range S_n, will in general depend upon n (although by this we do not mean to exclude the case in which μ_n is constant for all values of n), and perhaps will depend also on some further independent parameters, which we shall denote collectively by θ, with range Σ. We shall seek a variate $T = f(X)$, in which $f(X)$ is functionally independent of μ and of the parameters θ for μ on S_n, θ on Σ, and such that the distribution of $f(X) - f(\mu_n)$ tends as $n \to \infty$ to a normal distribution, while $\lim_{n\to\infty}\sigma_T^2 = c^2$, where c^2 is an absolute constant. It is implied here that in case the additional parameters θ are present, the function $f(X)$ may depend non-trivially on n; but if n is the only parameter on which the distribution of X depends, then $f(X)$ must be functionally independent of n.

A solution to the problem just proposed is given in certain cases by Theorems 3.1 and 3.2 below, which are suggested by the heuristic reasoning of the second paragraph of section 2.

THEOREM 3.1. *Let $\psi_n(x)$ be a non-negative function of x and n, defined almost everywhere and integrable*[4] *with respect to x over any finite interval of the x-axis for each $n > 0$. Let*

$$T = f(X) = \int_a^X \psi_n(x)\,dx,$$

where a is an arbitrary constant. Let $F_n(y)$ be the d.f. of the variate $Y = (X - \mu_n)\psi_n(\mu_n)$. Suppose further that a continuous d.f. $F(y)$ exists such that $\lim_{n\to\infty}F_n(y) = F(y)$ for all values of y. Then either one of the following two conditions is a sufficient condition for the d.f. $H_n(w)$ of the variate $W = f(X) - f(\mu_n)$ to tend uniformly to $F(w)$, $-\infty < w < \infty$:

(a) *To each w for which $0 < F(w) < 1$, there corresponds for all n sufficiently large at least one root $x = x_n$ to the equation*

$$\int_{\mu_n}^{x} \psi_n(u)\,du = w, \tag{3.1}$$

and this root x_n has the property that

$$\lim_{n\to\infty}(x_n - \mu_n)\psi_n(\mu_n) = w. \tag{3.2}$$

(b) *For all n sufficiently large, $\psi_n(\mu_n) > 0$, and $\lim_{n\to\infty}q_n(w) = 1$ uniformly in any closed finite subinterval of the open interval defined by $0 < F(w) < 1$, where*

$$q_n(w) = \frac{\psi_n(w[\psi_n(\mu_n)]^{-1} + \mu_n)}{\psi_n(\mu_n)}. \tag{3.3}$$

To prove this theorem we shall first suppose that condition (a) is satisfied. Let w_1 and w_2 be the end points of the open interval (possibly infinite) defined by $0 < F(w) < 1$. If w lies in this interval, and if n is large enough for the root x_n in (3.1) to exist, then from the monotonic character of $\int_{\mu_n}^{X} \psi_n(x)\,dx$ we can

[4] "Integrable" here means absolutely integrable in the sense of Lebesgue.

infer that

$$
\begin{aligned}
H_n(w) &= P[f(X) - f(\mu_n) \leqq w] = P\left[\int_{\mu_n}^{X} \psi_n(x)\,dx \leqq w\right] \\
&= P(X \leqq x_n) = P[Y \leqq (x_n - \mu_n)\psi_n(\mu_n)] \\
&= F_n[(x_n - \mu_n)\psi_n(\mu_n)].
\end{aligned} \tag{3.4}
$$

Since $F(w)$ is continuous, $\lim_{n\to\infty}F_n(w) = F(w)$ *uniformly* on any finite or infinite interval of values of w, as is well known.[5] Therefore $\lim_{n\to\infty}F_n(w_n) = F(w)$ if $\lim_{n\to\infty}w_n = w$. Thus from (3.2) and (3.4), we find that $\lim_{n\to\infty}H_n(w) = F(w)$ for $w_1 < w < w_2$.

If $w' \leqq w_1$, and $w_1 < w'' < w_2$, then $0 \leqq H_n(w') \leqq H_n(w'') = F(w'') + [H_n(w'') - F(w'')]$. We can make the right hand member of this relation less than any given positive number ϵ by first choosing w'' so that $F(w'') < \frac{1}{2}\epsilon$ (it will be remembered that $F(w)$ is a continuous d.f., and $F(w_1) = 0$) and then choosing n so large that the quantity in square brackets is also less than $\frac{1}{2}\epsilon$ in absolute value. Thus $\lim_{n\to\infty}H_n(w') = 0$. Similarly if $w' \geqq w_2$, we can show that $\lim_{n\to\infty}H_n(w') = 1$. Hence $\lim_{n\to\infty}H_n(w) = F(w)$ for all w, and it follows that the limit is uniform on any finite or infinite interval of values of w.

We shall now show that condition (a) in the theorem is a consequence of condition (b). The result follows at once from the following simple lemma:

LEMMA. *If $\gamma_n(w)$ is a non-negative function integrable over any finite interval of values of w; and if $\lim_{n\to\infty}\gamma_n(w) = 1$ uniformly in any finite closed subinterval of an interval $w_1 < w < w_2$, then for every value of w in this interval there exists for all n sufficiently large a solution $y = y_n$ of the equation $\int_0^y \gamma_n(z)\,dz = w$, and the solution y_n has the property that $\lim_{n\to\infty}y_n = w$.*

For it is clear that if w satisfies the inequality $w_1 < w < w_2$, and if $\eta > 0$ be chosen so that $w_1 < w - \eta < w + \eta < w_2$, then for all n sufficiently large,

$$\int_0^{w-\eta} \gamma_n(z)\,dz \leqq w \leqq \int_0^{w+\eta} \gamma_n(z)\,dz.$$

Thus for each n sufficiently large, there exists a root y_n of the equation $\int_0^y \gamma_n(z)\,dz = w$, and furthermore, this root satisfies the inequality $w - \eta \leqq y_n \leqq w + \eta$. Since η is arbitrarily small, the proof of the lemma is complete.

To apply the lemma, we make the change of variables $z = (u - \mu_n)\psi_n(\mu_n)$ in the integral in (3.1), which reduces it to the form

$$\int_0^y q_n(z)\,dz, \qquad y = (x - \mu_n)\psi_n(\mu_n), \tag{3.5}$$

and the conclusion that (a) is implied by (b) now follows at once.

[5] See [7], Theorem 11, pp. 29–30; also [8].

We add the remark that the uniformity of the limit of $q_n(z)$ in condition (b) may be replaced by the condition that for each closed finite sub-interval there exists a function $q(w)$ which dominates $q_n(w)$ for all n sufficiently large.

Our second theorem, which is stated in the terminology and notation of Theorem 3.1, is concerned with the limit of the variance of $T = f(X)$. From the mere fact that the distribution of W tends to a limiting form, it by no means follows that the mean and variance of the distribution of W approach those of the limiting form, as may be shown by trivial examples. Thus additional hypotheses on $\psi_n(x)$ and on the behavior of the distribution of Y become necessary.

THEOREM 3.2. *Let T (or $f(X)$), Y, $F_n(y)$ and $F(y)$ be defined as in Theorem 3.1. Let the mean and variance of the distribution defined by $F(y)$ exist and have respective values 0 and c^2. Then the following three conditions, taken together, are sufficient that*

$$\lim_{n\to\infty}[E(T) - f(\mu_n)] = 0, \tag{3.6}$$

$$\lim_{n\to\infty}\sigma_T^2 = c^2: \tag{3.7}$$

(i) *$E(Y^2)$ exists for $n > 0$, and $\lim_{n\to\infty}E(Y^2) = c^2$.*

(ii) *Condition (b) of Theorem 3.1 holds.*

(iii) *$f(Y[\psi_n(\mu_n)]^{-1} + \mu_n) - f(\mu_n) = O \mid Y \mid$ uniformly in n as $\mid Y \mid \to \infty$.*

As a preliminary step in the proof, we observe that (i) and the relations $\lim_{n\to\infty}F_n(y) = F(y)$, $c^2 = \int_{-\infty}^{+\infty} y^2\,dF(y)$, imply that the improper integral $\int_{-\infty}^{+\infty} y^2\,dF_n(y)$ converges uniformly in n for $n > 0$. As the integrand is positive, the following result is equivalent to the uniform convergence of the integral: *For every $\epsilon > 0$, there exist numbers A_1 and A_2, $A_1 < A_2$, such that for all n sufficiently large,*

$$\left(\int_{-\infty}^{A_1} + \int_{A_2}^{\infty}\right) y^2\,dF_n(y) < \epsilon.$$

To prove this, we write

$$\left(\int_{-\infty}^{A_1} + \int_{A_2}^{\infty}\right) y^2\,dF_n(y) = [E(Y^2) - c^2] + \left[\int_{A_1}^{A_2} y^2\,dF(y) - \int_{A_1}^{A_2} y^2\,dF_n(y)\right] + \left[c^2 - \int_{A_1}^{A_2} y^2\,dF(y)\right].$$

We first choose A_1 and A_2 so that the last bracket here is less than $\frac{1}{2}\epsilon$ in absolute value. By condition (i), the first bracket approaches zero as n tends to infinity, and the Helly-Bray theorem [10, p. 15] states that the second bracket also approaches zero as n tends to infinity, so for all n sufficiently large, the sum of the first two brackets is in absolute value less than $\frac{1}{2}\epsilon$.

It is important to notice that we can always choose A_1 and A_2 in the above

demonstration so that $A_1 > w_1$, $A_2 < w_2$, where w_1 and w_2 are as usual the endpoints of the interval defined by $0 < F(w) < 1$.

To continue with the proof of the theorem, we remark that by a change of variables similar to the one used to derive (3.5), the function $W = f(X) - f(\mu_n)$ may be expressed as a function of Y in the following manner:

$$W = \int_{\mu_n}^{X} \psi_n(x)\,dx = \int_0^Y q_n(w)\,dw = Q_n(Y),$$

where $q_n(w)$ is given by (3.3). In terms of W, (3.6) and (3.7) become, respectively,

$$\lim_{n\to\infty} E(W) = 0, \tag{3.8}$$

$$\lim_{n\to\infty} \{E(W^2) - [E(W)]^2\} = c^2, \tag{3.9}$$

and these are the equations which we now establish.

Conditions (ii) and (iii) obviously imply that $\lim_{n\to\infty} Q_n(y) = y$ uniformly in any finite closed subinterval of the interval $w_1 < y < w_2$, and that a constant M exists such that $|\,Q_n(y)| \leqq M\,|\,y\,|$ for all n. If $E(Y^2)$ exists, so will $E(Y)$. Now

$$\begin{aligned} E(W) &= \int_{-\infty}^{+\infty} Q_n(y)\,dF_n(y) \\ &= \int_{-\infty}^{+\infty} Q_n(y)\,dF_n(y) - \int_{-\infty}^{+\infty} y\,dF_n(y) \\ &= \left(\int_{-\infty}^{A_1} + \int_{A_2}^{\infty}\right)[Q_n(y) - y]\,dF_n(y) + \int_{A_1}^{A_2} [Q_n(y) - y]\,dF_n(y), \end{aligned}$$

where $w_1 < A_1 < A_2 < w_2$. Therefore

$$|\,E(W)\,| \leqq \left(\int_{-\infty}^{A_1} + \int_{A_2}^{\infty}\right)(M+1)\,|\,y\,|\,dF_n(y) + \int_{A_1}^{A_2} |\,Q_n(y) - y\,|\,dF_n(y).$$

From the uniform convergence of $\int_{-\infty}^{+\infty} y^2 dF_n(y)$, proved above, we can conclude that the pair of improper integrals in this inequality can be made less than an arbitrary $\frac{1}{2}\epsilon > 0$ by proper choice of A_1 and A_2. The third integral approaches zero, by the general Helly-Bray Theorem [10, p. 16], and so becomes less than $\frac{1}{2}\epsilon$ for all n sufficiently large. Thus we have established (3.8). To show that (3.9) is true, we have merely to prove that $\lim_{n\to\infty} E(W^2) = c^2$. Since $E(Y^2) = \int_{-\infty}^{+\infty} y^2 dF_n(y)$, we may write

$$E(W^2) - c^2 = \int_{-\infty}^{+\infty} \{[Q_n(y)]^2 - y^2\}\,dF_n(y) + [E(Y^2) - c^2].$$

The integral may be shown to approach zero by the argument used in the case of $E(W)$, and the required result then follows from condition (i) of the theorem. The proof is now complete.

The sufficient conditions in Theorem 3.2 can be modified in various more or less obvious ways. The existence of the limiting d.f. $F(y)$ was essentially used in the proof only to secure the uniform convergence of $\int_{-\infty}^{+\infty} y^2 dF_n(y)$. Condition (ii) can again be modified along the lines suggested at the end of the proof of Theorem 3.1. Condition (iii) was used only to secure the uniform convergence of the integral $\int_{-\infty}^{+\infty} [Q_n(y)]^2 dF_n(y)$.

For later reference, we shall supplement Theorems 3.1 and 3.2 with the following simple result, which is practically self-evident.

THEOREM 3.3. *Let the distribution of a variate Y depend upon a parameter n, let $F_n(y)$ be the d.f. of Y, and let $F(y)$ be a continuous d.f. with the property that $\lim_{n\to\infty} F_n(y) = F(y)$. Let a_n be a function of n such that $\lim_{n\to\infty} a_n = a \neq 0$. Then the d.f. of the variate $Z = a_n Y$ tends as $n \to \infty$ to the d.f. $F(z/a)$ if $a > 0$, and to the d.f. $1 - F(z/a)$ if $a < 0$. If the variance of Y exists and tends to c^2 as $n \to \infty$, then the variance of $a_n Y$ tends to a^2c^2 as $n \to \infty$.*

If $F(y)$ is the d.f. of a reduced normal distribution, i.e.,

$$F(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}t^2}\, dt,$$

then $F(z/a)$ is also the d.f. of a normal distribution with mean zero and variance a^2. More generally, any affine transformation of a normal variate yields another normal variate.

4. Applications. The theorems of the preceding section have the effect of referring the properties of the distribution of the transformation $T = f(X)$ of Theorem 3.1 back to those of the distribution of a related variate Y. In the applications given in the present section, we shall let $\psi_n(\mu_n)$ be proportional to the reciprocal of the standard deviation of X. The theorems of section 3 state in this case that if the reduced, or standardized, distribution of X approaches a limiting form, then under certain circumstances, the distribution of $f(X) - f(\mu_n)$ will approach a similar limiting form, and σ_T^2 will approach a quantity independent at least of n. In the applications considered here, the reduced distribution of X will always approach the reduced normal distribution.

(I) **The square root transformation for a variate with a Poisson exponential distribution.** *Let X have a Poisson exponential distribution with parameter n. If α is an arbitrary constant, and if*

$$T = f(X) = \begin{cases} \sqrt{X + \alpha}, & X \geqq -\alpha \\ 0, & X < -\alpha \end{cases} \tag{4.1}$$

then the distribution of $T - \sqrt{n+\alpha}$ *tends as* $n \to \infty$ *to a normal distribution which has mean zero and variance* $\frac{1}{4}$, *and* $\lim_{n\to\infty}\sigma_T^2 = \frac{1}{4}$. For $\mu_n = n$, $\sigma_X = \sqrt{n}$, and it is well known[6] that the distribution of the reduced variate $(X-n)/\sqrt{n}$ tends to the reduced normal distribution as $n \to \infty$. By Theorem 3.3, the distribution of the variate

$$Y = \frac{X-n}{2\sqrt{n+\alpha}} = \frac{1}{2}\cdot\sqrt{\frac{n}{n+\alpha}}\cdot\frac{X-n}{\sqrt{n}},$$

will tend to normality as $n \to \infty$, and the variance of Y will tend to the value $\frac{1}{4}$, which is also the variance of the limiting distribution. Setting

$$\psi_n(x) = \begin{cases} \dfrac{1}{2\sqrt{x+\alpha}}, & x > -\alpha \\ 0, & x \leqq -\alpha, \end{cases}$$

we obtain from $T = f(X) = \int_a^X \psi_n(x)dx$ the formula given in (4.1). To prove the statement in italics, we must show that conditions (ii) and (iii) of Theorem 3.2 are satisfied. We have, assuming $n > -\alpha$,

$$q_n(w) = \begin{cases} \left(1 + \dfrac{2w}{\sqrt{n+\alpha}}\right)^{-1}, & w > -\frac{1}{2}\sqrt{n+\alpha} \\ 0, & w \leqq -\frac{1}{2}\sqrt{n+\alpha}, \end{cases}$$

so clearly (ii) is satisfied. Also,

$$W = f(Y[\psi_n(\mu_n)]^{-1} + \mu_n) - f(\mu_n)$$
$$= \begin{cases} \sqrt{2Y\sqrt{n+\alpha} + n + \alpha} - \sqrt{n+\alpha}, & Y > -\frac{1}{2}\sqrt{n+\alpha} \\ -\sqrt{n+\alpha}, & Y \leqq -\frac{1}{2}\sqrt{n+\alpha}, \end{cases}$$

from which it follows at once that $|W| < 2|Y|$ for all Y, and so (iii) is satisfied.

The degree of approximation involved in the equation $\lim_{n\to\infty}\sigma_T^2 = \frac{1}{4}$ has been investigated numerically by Bartlett [1] for values of n from .5 to 15.0 in the cases $\alpha = 0$ and $\alpha = \frac{1}{2}$. He found that the variance of $\sqrt{X + (\frac{1}{2})}$ is considerably closer to the limit $(\frac{1}{4})$ for $1 \leqq n \leqq 10$ than is the variance of $\sqrt{X}$. At $n = 15$, the variance of $\sqrt{X}$ is .256, and that of $\sqrt{X + (\frac{1}{2})}$ is .248.

The question of the degree of convergence to normality and of the possibility of selecting an optimum value of α remain open. By expanding the function $\sqrt{X+\alpha}$ in a Taylor series about $X = n$ with remainder in the form due to Schlömilch, it is possible to derive as accurate an estimate of $|\sigma_T^2 - (\frac{1}{4})|$ as may

[6] See (e.g.) [9].

be desired. A rough result easily obtainable by this method is that $|\sigma_T^2 - (\frac{1}{4})| \leqq 3/(4n)$, $n > 0$.

(II) **The square root transformation for a variate with a Γ distribution.** *Let X have a distribution whose density function is of the following type:*

$$\varphi(x) = \begin{cases} 0 & x \leqq 0 \\ Kx^{\frac{1}{2}n-1}e^{-hx}, & x \geqq 0,\ h > 0. \end{cases} \tag{4.2}$$

If α is an arbitrary constant, and if

$$T = f(X) = \begin{cases} \sqrt{X + \alpha}, & X \geqq -\alpha \\ 0\ , & X < -\alpha, \end{cases} \tag{4.3}$$

then the distribution of $T - \sqrt{(n/2h) + \alpha}$ tends as $n \to \infty$ to a normal distribution which has mean zero and variance $1/4h$, and $\lim_{n\to\infty}\sigma_T^2 = 1/(4h)$. For $\mu_n = n/(2h)$, $\sigma_X = \sqrt{n}/(h\sqrt{2}) = \sqrt{\mu_n/h}$. The distribution of the reduced variate tends to normality as $n \to \infty$,[7] so that of the variate

$$Y = \frac{x - \mu_n}{2\sqrt{\mu_n + \alpha}} = \frac{1}{2}\sqrt{\frac{n}{nh + 2h^2\alpha}} \cdot \frac{x - \mu_n}{\sqrt{\mu_n/h}}$$

tends to normality also with limiting variance $1/(4h)$. Setting

$$\psi_n(x) = \begin{cases} \dfrac{1}{2\sqrt{x + \alpha}}, & x > -\alpha \\ 0\ , & x \leqq -\alpha, \end{cases}$$

we obtain T in (4.3) from the relation $T = \int_{-\alpha}^{x} \psi_n(x)\,dx$. The work of verifying that the conditions of Theorem 3.2 are satisfied is the same as in the case of the Poisson exponential distribution treated above, and will not be repeated.

For example, if s^2 denotes the variance of a random sample of $n + 1$ observations drawn from a normal parent distribution with variance σ^2, then it is well known that $(n + 1)s^2$ is distributed according to (4.2) with $h = 1/(2\sigma^2)$. We thus can deduce the further facts, also well known, that the distribution of $\sqrt{n + 1}\, s - \sigma\sqrt{n}$ tends to normality, and that the variance of $s\sqrt{n + 1}$ approaches the limiting value $\frac{1}{2}\sigma^2$. If n is an integer and $h = \frac{1}{2}$, the distribution defined by (4.2) is called a χ^2 distribution with n degrees of freedom, and the variate is often denoted by χ^2. Our conclusion in this case is that the distribution of $\sqrt{2\chi^2} - \sqrt{2n}$ tends to a normal one with zero mean and unit variance. From this result and the fact that $\sqrt{2\,n-1} - \sqrt{2n} = O(n^{-\frac{1}{2}})$, it follows immediately that $\sqrt{2\chi^2} - \sqrt{2n - 1}$ has the same limiting distribution as $\sqrt{2\chi^2} - \sqrt{2n}$. This result,[8] due to Fisher, is familiar to all users of his table of the probability levels of χ^2.

[7] See (e.g.) [9].

[8] For a discussion of the degree of convergence involved here, see [9].

(III) **The inverse sine transformation for a binomial variate.** *Let X have a binomial relative frequency distribution with parameter p and the n values $0, 1/n, 2/n, \cdots, n/n$. If α is an arbitrary constant, and if*

$$T = f(X) = \begin{cases} \sqrt{n}\sin^{-1}\sqrt{X + \frac{\alpha}{n}}, & -\frac{\alpha}{n} \leqq X \leqq 1 - \frac{\alpha}{n} \\ 0, & X < -\frac{\alpha}{n}, \quad X > 1 - \frac{\alpha}{n}, \end{cases} \tag{4.4}$$

where T is measured in radians, then the distribution of $T - \sqrt{n}\sin^{-1}\sqrt{p + (\alpha/n)}$ tends as $n \to \infty$ to a normal distribution which has mean zero and variance $\frac{1}{4}$, and $\lim_{n\to\infty}\sigma_T^2 = \frac{1}{4}$. For here, $\mu_n = p$, and $\sigma_X^2 = pq/n$, where $q = 1 - p$; and the familiar DeMoivre-Laplace theorem states that the distribution of the reduced variate $\sqrt{n}(X - p)/\sqrt{pq}$ will tend to normality as $n \to \infty$. Hence by Theorem 3.3 the distribution of

$$Y = \frac{\sqrt{n}(X - p)}{2\sqrt{\left(p + \frac{\alpha}{n}\right)\left(q - \frac{\alpha}{n}\right)}} \tag{4.5}$$

will tend to normality with a limiting variance of $\frac{1}{4}$, which is also the variance of the limiting distribution. Setting

$$\psi_n(x) = \begin{cases} \dfrac{\sqrt{n}}{2\sqrt{\left(x + \frac{\alpha}{n}\right)\left(1 - x - \frac{\alpha}{n}\right)}}, & -\frac{\alpha}{n} < x < 1 - \frac{\alpha}{n} \\ 0 & x \leqq -\frac{\alpha}{n}, x \geqq 1 - \frac{\alpha}{n}, \end{cases}$$

we obtain (4.4) from the integral

$$T = \int_{\alpha/n}^{x} \psi_n(x)\,dx.$$

In proving the conditions (ii) and (iii) of Theorem 3.2 are satisfied, we shall assume for simplicity that $\alpha = 0$. We find that

$$q_n(w) = \begin{cases} \left(1 + 2w\dfrac{q - p}{\sqrt{npq}} - \dfrac{4w^2}{n}\right)^{-1}, & -\frac{1}{2}\sqrt{\frac{np}{q}} < w < \frac{1}{2}\sqrt{\frac{nq}{p}} \\ 0, & w \leqq -\frac{1}{2}\sqrt{\frac{np}{q}}, \quad w \geqq \frac{1}{2}\sqrt{\frac{nq}{p}}, \end{cases}$$

so obviously (ii) is satisfied. From the Law of the Mean in the form due to Schlömilch, we have

$$\begin{aligned} W &= \sqrt{n}\sin^{-1}\sqrt{p + 2\sqrt{\frac{pq}{n}}\,Y} - \sqrt{n}\sin^{-1}\sqrt{p} \\ &= 2Y\left[\frac{1 - \theta}{\left(1 + 2\theta\sqrt{\frac{q}{np}}\,Y\right)\left(1 - 2\theta\sqrt{\frac{p}{nq}}\,Y\right)}\right]^{\frac{1}{2}}, \end{aligned} \tag{4.6}$$

$$0 < \theta < 1, \qquad -\frac{1}{2}\sqrt{\frac{np}{q}} < Y < \frac{1}{2}\sqrt{\frac{nq}{p}}.$$

The denominator of the coefficient of $2Y$ here is a quadratic function of Y with a negative coefficient of Y^2, and so must assume its least value in the Y range indicated in (4.6) at one end or the other of the range. From this it is readily seen that the coefficient of $2Y$ is actually always less than unity. For values of Y outside the range, the second member of (4.6) indicates that $W = O(\sqrt{n}) = O(Y)$. Hence (iii) is satisfied, and the proof of the statement in italics is complete for the case $\alpha = 0$. The more general case presents no important new difficulties.

In practice, it is often convenient to express X as a percentage. This merely has the effect of multiplying Y in (4.5) by 100. We find in this case that $\sqrt{n}\ \sin^{-1}\sqrt{X + 100\alpha/n} - \sqrt{n}\ \sin^{-1}\sqrt{100p + 100\alpha/n}$ has a distribution approaching normality, and $\sigma_T \to 50$ instead of $\frac{1}{2}$.

Bartlett [1] gives numerical results in the cases $n = 10$, $\alpha = 0$ and $n = 10$, $\alpha = \frac{1}{2}$, which indicate that perhaps the choice $\alpha = \frac{1}{2}$ is more suitable if the estimated p is near 0 or 1, but the choice $\alpha = 0$ is preferable if the estimated p lies between .3 and .7. However, there seems to be no good reason to believe that these conclusions should be valid for other values of n. The question of an optimum α, and of the degree of convergence to normality remain open. We note in passing that the latter problem could doubtless be profitably studied by combining the methods of proof of Theorem 3.1 with the results of Uspensky [15, pp. 129–130] on the degree of approximation of the reduced binomial d.f. to the normal d.f.

IV. **Other transformations of a binomial variate.** *Let X have a binomial relative frequency distribution with the parameter p and the n values $0, 1/n, 2/n, \cdots, n/n$.*

(a) *If*

$$T = f(X) = \begin{cases} \sqrt{n}\ \sinh^{-1}\sqrt{X} = \sqrt{n}\ \log\left(\sqrt{X} + \sqrt{1+X}\right),^{9} & X \geqq 0 \\ 0 \quad , & X < 0, \end{cases}$$

then the distribution of $T - \sqrt{n}\ \sinh^{-1}\sqrt{p}$ tends as $n \to \infty$ to a normal distribution which has mean zero and variance $q/(4 + 4p)$, and $\lim_{n\to\infty}\sigma_T^2 = q/(4 + 4p)$.

(b) *If*

$$T = f(X) = \begin{cases} \sqrt{n}\ \log X, & X > 0, \\ 0 \quad , & X \leqq 0, \end{cases}$$

then the distribution of $T - \sqrt{n}\ \log p$ tends as $n \to \infty$ to a normal distribution which has mean zero and variance q/p, and $\lim_{n\to\infty}\sigma_T^2 = q/p$.

(c) *If*

$$T = f(X) = \begin{cases} \frac{1}{2}\sqrt{n}\ \log\dfrac{X}{1-X}, & 0 < X < 1, \\ 0 \quad , & X \leqq 0, \quad X \geqq 1, \end{cases}$$

[9] All logarithms in this paper are to the base e.

then the distribution of $T - \frac{1}{2}\sqrt{n}\log\frac{p}{1-p}$ *tends as* $n \to \infty$ *to a normal distribution which has mean zero and variance* $1/(4pq)$, *and* $\lim_{n\to\infty}\sigma_T^2 = 1/(4pq)$.

Since the limiting variance of each of these transformations involves the parameter p, they are not to be regarded as solutions of the problem of asymptotic variance stabilization proposed at the beginning of section 3, although it is perhaps of some interest that their distributions become asymptotically normal.

In case (a), $f'(x) = \sqrt{n}/(2\sqrt{x^2+x})$, $x > 0$. Setting $\psi_n(x) = f'(x)$, $x > 0$, and $\psi_n(x) = 0$, $x \leqq 0$, we obtain

$$Y = (X-p)\psi_n(p) = \frac{\sqrt{n}\,(X-p)}{\sqrt{pq}}\cdot\frac{\sqrt{q}}{2\sqrt{1+p}}, \tag{4.7}$$

and this variate obviously has the limiting distribution ascribed to $T - \sqrt{n}\sinh^{-1}\sqrt{p}$ in the statement in italics. The truth of that statement now follows by an argument similar to that used in the case of the inverse sine transformation.

If p is allowed to vary with n in such a way that $\lim_{n\to\infty} np = \infty$, it is known that the reduced distribution of X will still tend to normality.[10] If we suppose that $\lim_{n\to\infty} p = 0$, but $\lim_{n\to\infty} np = \infty$, we find from Theorem 3.3 that the limiting distribution of Y in (4.7) will be normal with mean zero and variance $\frac{1}{4}$, and that $\sigma_Y^2 \to \frac{1}{4}$. It is easily verified that the conditions (ii) and (iii) of Theorem 3.2 are still satisfied, so we find that the limiting distribution of $[\sqrt{n}\sinh^{-1}\sqrt{X} - \sqrt{n}\sinh^{-1}\sqrt{p}]$ is normal, with mean zero and variance $\frac{1}{4}$, and $\sigma_T^2 \to \frac{1}{4}$. However, since n is now the only independent parameter, we cannot here regard the transformation $T = \sqrt{n}\sinh^{-1}\sqrt{X}$ as a solution of the problem of variance stabilization, because the variate T depends explicitly upon n.

If in case (b) we proceed as in case (a), we obtain as the analogue of (4.7) the formula

$$Y = (X-p)\psi_n(p) = \frac{\sqrt{n}\,(X-p)}{\sqrt{pq}}\sqrt{\frac{q}{p}},$$

and this variate has the limiting distribution ascribed to $T - \sqrt{n}\log X$ in the statement in italics. It now turns out that although condition (ii) of Theorem 3.2 is satisfied, condition (iii) is not satisfied. We are then faced with the problem of proving directly that the improper integral

$$\int_{-\sqrt{n}}^{+\infty} [\sqrt{n}\log(p + py/\sqrt{n}) - \sqrt{n}\log p]^2\, dF_n(y)$$

converges uniformly.[11] The trouble occurs only at the lower limit of integration, and may be resolved by first integrating by parts, then dividing the range

[10] See (e.g.) [9].

[11] See the remarks following the proof of Theorem 3.2.

$(-\sqrt{n}, A_1)$ into two ranges $(-\sqrt{n}, -\log n)$ and $(-\log n, A_1)$, and then applying Uspensky's results [15, pp. 129–130], on the degree of approximation involved in the DeMoivre-Laplace theorem.

Case (c) may be handled in a similar manner.

5. The logarithmic transformation. We shall suppose throughout this section that X is a variate whose mean μ_n and standard deviation σ in the relation $\sigma = k_n(\mu_n + \alpha)$, where α is an arbitrary constant, $k_n > 0$, and $\lim_{n\to\infty} k_n$ exists and is finite. If k_n is constant for all n, say $k_n = k > 0$, and if we use the heuristic argument of the second paragraph of section 2 to attempt to find a transformation which will stabilize the variance of X at k^2, we arrive at the function $T = \log(X + \alpha)$, $X > -\alpha$. It is the purpose of this section to study the asymptotic properties of this transformation.

The theory of such a transformation differs in certain important respects from that of the transformations considered in sections 3 and 4. For one thing, our starting point in the study of each transformation considered in section 4 was the fact that although $P(X < 0) = 0$, nevertheless the reduced distribution of X tended to normality as $n \to \infty$. But in the present case, if X is a variate such that $P(X \leqq -\alpha) = 0$, then the corresponding reduced variate $Y = (X - \mu_n)/[k_n(\mu_n + \alpha)]$ has a d.f. $F_n(y)$ such that $F_n(-1/k_n) = 0$. Thus if $\lim_{n\to\infty} k_n = k > 0$, the limiting distribution of Y, if it exists, must have a d.f. $F(y)$ such that $F(-1/k - 0) = 0$. Therefore the limiting distribution of Y can never be normal if $k > 0$.

Moreover (in contrast to the situation in Theorem 3.1) if the reduced variate Y does have a limiting distribution, the variate

$$(5.1)\quad W = \frac{1}{k_n}\log(X + \alpha) - \frac{1}{k_n}\log(\mu_n + \alpha) = \int_{\mu_n}^{X} \frac{1}{k_n(u + \alpha)}\,du, \quad X > -\alpha$$

may have a limiting distribution which is not the same as that of Y. More specifically, we have the following result:

THEOREM 5.1. *Let* $P(X \leqq -\alpha) = 0$, *let* $\lim_{n\to\infty} k_n = k \geqq 0$, *let* $F_n(y)$ *be the d.f. of the reduced variate*

$$Y = \frac{X - \mu_n}{k_n(\mu_n + \alpha)},$$

and let $H_n(w)$ *be the d.f. of the variate* W *given by* (5.1). *If a continuous d.f.* $F(y)$ *exists such that* $\lim_{n\to\infty} F_n(y) = F(y)$ *for all* y, *then*

$$\lim_{n\to\infty} H_n(w) = \begin{cases} F\left[\dfrac{e^{kw} - 1}{k}\right], & k > 0 \\ F(w), & k = 0. \end{cases}$$

The proof is simpler than the statement; essentially we have only to notice that

$$H_n(w) = P\left[-\frac{1}{k_n} < Y \leqq \frac{e^{k_n w} - 1}{k_n}\right]$$
$$= F_n\left[\frac{e^{k_n w} - 1}{k_n}\right], \qquad -\infty < w < \infty,$$

and apply the reasoning used above in connection with (3.4).

From the study of the distribution of T, we now turn for a moment to the question of the limit if σ_T^2. Here the situation is more consistent with the results of section 3.

THEOREM 5.2. *Under the hypotheses of Theorem* 5.1 *and under the additional conditions that the improper integral* $\int_{-\infty}^{0} w^2\, dH_n(w)$ $\left(\text{or } \int_{-1/k_n}^{0} k_n^{-2}[\log(1 + k_n y)]^2\, dF_n(y)\right)$ *converges uniformly in n and that* $\int_{-\infty}^{+\infty} y^2\, dF(y) = 1 = E(Y^2)$, *the following relations hold:*

$$\lim_{n\to\infty} E(W) = \begin{cases} \int_{-1/k}^{\infty} \frac{1}{k} \log(1 + ky)\, dF(y), & k > 0, \\ 0, & k = 0, \end{cases} \tag{5.2}$$

$$\lim_{n\to\infty} E(W^2) = \begin{cases} \int_{-1/k}^{\infty} \frac{1}{k^2} [\log(1 + ky)]^2\, dF(y), & k > 0 \\ 1, & k = 0 \end{cases} \tag{5.4}$$

The variance σ_T^2 of the variate $T = \log(X + \alpha)$ is related to these mean values by the equation $\sigma_T^2 = k_n^2\{E(W^2) - [E(W)]^2\}$. Thus if $F(y)$ is independent of any unknown parameters θ, and if k is positive and is presumed to have the same value for all variates in any given problem, then the transformation $T = \log(X + \alpha)$ is seen to yield an asymptotic stabilization of the variance under the conditions of Theorem 5.2. If $k = 0$, we find from either Theorem 5.2 or the proof of Theorem 5.2 that $T = \log(X + \alpha)$ converges stochastically to $\log(\mu_n + \alpha)$.

The proof of Theorem 5.2 is similar to that of Theorem 3.2 and will be omitted.

Theorem 5.1 raises the following question: Just what limiting distribution must Y have if $k > 0$, in order that the distribution of W tend to normality? To answer this, we shall note the following simple non-asymptotic result:

THEOREM 5.3. *A necessary and sufficient condition that X have a continuous distribution with density function*

$$\varphi(x) = \begin{cases} \frac{1}{\sqrt{2\pi \log(k^2 + 1)}} \frac{1}{x + \alpha} \\ \qquad \times \exp\left[\frac{-\left(\log \frac{(x + \alpha)\sqrt{k^2 + 1}}{u + \alpha}\right)^2}{2 \log(k^2 + 1)}\right], & x > -\alpha \\ 0, & x \leqq -\alpha \end{cases} \tag{5.4}$$

for which $\sigma_X = k(\mu + \alpha)$, *is that the variate* $T = \log(X + \alpha)$ *have a normal distribution with mean* $\log(\mu + \alpha) - \log\sqrt{k^2 + 1}$ *and variance* $\log(k^2 + 1)$.

The proof may be given by a routine change of variables.[12] It is to be noticed that the heuristic argument of the second paragraph of section 2 would lead to the incorrect result that the variance of T was k^2 instead of $\log(k^2 + 1)$. In case $k = 1$, the mean and variance of T are respectively $\log(\mu + \alpha) - .347$ and .693. If the transformation $T = \log_{10}(X + \alpha)$ is used, the new mean is $\log_{10}(\mu + \alpha) - \log_{10}\sqrt{k^2 + 1}$ and the new variance is $.189 \log(k^2 + 1$, which for values of k near zero has the approximate value $.189k^2$.[13]

If X is distributed according to (5.4), the density function $F'(y)$ of the corresponding reduced variate $Y = (X - \mu)/[k(\mu + \alpha)]$ is

$$(5.5)\quad F'(y) = \begin{cases} \dfrac{k}{\sqrt{2\pi \log(k^2 + 1)}} \cdot \dfrac{1}{1 + ky} \\ \qquad \times \exp\left[-\dfrac{\{\log[(1 + ky)\sqrt{k^2 + 1}]\}^2}{2 \log(k^2 + 1)}\right], & y > -\dfrac{1}{k} \\ 0 & y \leqq -\dfrac{1}{k}. \end{cases}$$

The d.f. of the variate $W = k^{-1}[\log(X + \alpha) - \log(\mu + \alpha)]$ is $F[(e^{kw} - 1)/k]$, and, of course, the distribution of W is normal with mean $-k^{-1} \log\sqrt{k^2 + 1}$, and variance $k^{-2} \log(k^2 + 1)$. These are the respective values of the integrals in (5.2) and (5.3).

If now the distribution of X depends on a parameter n in such a way that as $n \to \infty$, the distribution of the corresponding reduced variate $Y = (X - \mu_n)/[k_n(\mu_n + \alpha)]$ tends to the distribution given by (5.5), it follows from the above remarks and from Theorem 5.1 that the variate W given by (5.1) has a normal limiting distribution. Furthermore, under the uniform convergence condition of Theorem 5.2, it follows that σ_T^2 tends to the value $\log(k^2 + 1)$, where $T = \log(X + \alpha)$.

These facts provide a sound mathematical basis for the use of the logarithmic transformation, which has had a long history of empirical success in problems of normalization [12, chapter XVI] and stabilization ([6], [16]). When it appears from a reasonably large number of observations on a variate (which is essentially bounded from below) that the standard deviation of the variate is proportional to the mean, then a possible specification for the variate is a distribution of the form (5.4); or, at least for large values of μ, it may be assumed that the distribution of the reduced variate is given by (5.5). Then the variate $T = \log(X + \alpha)$, where $-\alpha$ is any number less than the lower bound of X, will be exactly or approximately normally distributed with a variance independent of the value of μ.

Since (5.4) is only one of an infinity of various different types of distribution

---

[12] Finney [11] has considered the problem of efficiently estimating the variance of the X of Theorem 5.3 in the case $\alpha = 0$. (The actual density function (5.4) appears nowhere in his paper.)

[13] Given (without explanation) by Cochran [6, p. 165].

in which the mean and standard deviation are proportional, the user of a logarithmic transformation in the analysis of variance should always apply tests for departure from normality to the observed distribution of T values. From the point of view of specification, the situation here would seem to be less reassuring than in the cases considered in section 4. While it is true that the Poisson exponential distribution is only one of many types of distribution in which the variance and mean are equal, nevertheless the specification of a Poisson distribution can generally be preceded by a fairly strong chain of *a priori* inductive reasoning. This would not seem to be the case in the specification of (5.4). Theorems 5.1 and 5.2 furnish some grounds for a suspicion that the logarithmic transformation may possibly be more successful in stabilizing the variance than in normalizing the data. The burden of proof, however, lies with the experimenter.[14]

REFERENCES

[1] M. S. Bartlett, "The square root transformation in the analysis of variance," *Jour. Roy. Stat. Soc. Suppl.*, Vol. 3 (1936), pp. 68–78.

[2] Geoffrey Beall, "The transformation of data from entomological field experiments so that the analysis of variance becomes applicable," *Biometrika*, Vol. 32 (1942), pp. 243–262.

[3] C. I. Bliss, "The transformation of percentages for use in the analysis of variance," *Ohio Jour. Science*, Vol. 38 (1938), pp. 9–12.

[4] A. Clark and W. H. Leonard, "The analysis of variance with special reference to data expressed as percentages," *Jour. Amer. Soc. Agron.*, Vol. 31 (1939), pp. 55–56.

[5] W. G. Cochran, "The analysis of variance when experimental errors follow the Poisson or binomial laws," *Annals of Math. Stat.*, Vol. 9 (1940), pp. 335–347.

[6] W. G. Cochran, "Some difficulties in the statistical analysis of replicated experiments," *Empire Jour. Expt. Agric.*, Vol. 6 (1938), pp. 157–175.

[7] H. Cramér, *Random Variables and Probability Distributions*, Cambridge, 1937.

[8] J. H. Curtiss, "A note on the theory of moment generating functions," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 430–433.

[9] J. H. Curtiss, "Convergent sequences of probability distributions," *Am. Math. Monthly*, Vol. 50 (1943), pp. 94–105.

[10] G. C. Evans, *The Logarithmic Potential*, New York, 1927.

[11] D. J. Finney, "On the distribution of a variate whose logarithm is normally distributed," *Jour. Roy. Stat. Soc. Suppl.*, Vol. 7 (1941), pp. 155–161.

[12] Arne Fisher, *The Mathematical Theory of Probabilities*, Second edition, New York, 1930.

[13] R. A. Fisher and F. Yates, *Statistical Tables*, London, 1938.

[14] L. H. C. Tippett, "Statistical methods in textile research. Part 2, Uses of the binomial and Poisson distributions," *Shirley Inst. Mem.*, Vol. 13 (1934), pp. 35–72.

[15] J. V. Uspensky, *Introduction to Mathematical Probability*, New York, 1937.

[16] C. B. Williams, "The use of logarithms in the interpretation of certain entymological problems," *Annals of Appl. Biol.*, Vol. 24 (1937), pp. 404–414.

[14] A transformation closely related to the logarithmic one is $T = k^{-1} \sinh^{-1}(kX)^{\frac{1}{2}}$, where k is an estimate of the Charlier coefficient of disturbancy of a Poisson distribution. This transformation has recently been studied from an empirical point of view by Beall [2]; it was suggested by the heuristic argument of section 2 applied to the case in which $\sigma^2 = \mu + k\mu^2$. Beall presents evidence that for the particular data which he considered, the transformation seemed to stabilize the variance and normalize. A mathematical theory would follow the lines laid down above in the case of $T = \log(X + \alpha)$.

ON FUNDAMENTAL SYSTEMS OF PROBABILITIES OF A FINITE NUMBER OF EVENTS

By Kai Lai Chung

Tsing Hua University, Kunming, China

We consider a probability function $P(E)$ defined over the Borel set of events generated by the n arbitrary events $E_1, \cdots, E_n$, which will be denoted by $\mathfrak{L}(1, \cdots, n)$.

We use the same notations as in the author's former paper[1], with the following abbreviations. We denote a combination $(\alpha_1 \cdots \alpha_a)$ simply by (α), and use the corresponding Latin letter a for its number of members. Similarly we write (β) for $(\beta_1 \cdots \beta_b)$, but (ν) for $(1, \cdots, n)$. We say that (β) belongs to (α) and write $(\beta) \in (\alpha)$ when and only when the set $(\beta_1 \cdots \beta_b)$ is a subset of $(\alpha_1 \cdots \alpha_a)$. Then and then only we write $(\alpha) - (\beta)$ for the subset of elements of (α) that do not belong to (β); thus we may write it as (γ) with $c = a - b$. When and only when (α) and (β) have no common elements, we write $(\alpha) + (\beta)$ for the set of elements that belong either to (α) or to (β); thus we may write it as (γ), with $c = a + b \leqq n$. We note the case for empty sets: $(0) + (0) = (0)$. Now we can write $p_{[(\alpha)]}$ for $p_{[\alpha_1 \cdots \alpha_a]}$, $p_{((\alpha))}$ for $p_{\alpha_1 \cdots \alpha_a}$, $p_b((\alpha))$ for $p_b(\alpha_1 \cdots \alpha_a)$, etc. Further we denote by $p_{[b]}((\alpha))$ $(1 \leqq b \leqq a \leqq n)$ the probability of the occurrence of exactly b events out of $E_{\alpha_1}, \cdots, E_{\alpha_a}$, and write

$$P_a^{(m)}((\nu)) = \sum_{(\alpha) \in (\nu)} p_m((\alpha)), \qquad P_a^{[m]}((\nu)) = \sum_{(\alpha) \in (\nu)} p_{[m]}((\alpha));$$

since a is fixed by the left-hand sides, the summations on the right-hand sides are to be extended to all the $\binom{n}{a}$-combinations of (ν).

A sum written $\sum_{(\beta) \in (\alpha)}$ is to be extended to all combinations (β), $b = 0, 1, \cdots, a$ belonging to (α), when b is not previously fixed; it is to be extended to all the $\binom{a}{b}$-combinations belonging to (α), when b is previously fixed.

Definition 1. *A system of quantities is said to form a fundamental system of probabilities for a set of events if and only if the probability of every event in the set can be expressed in terms of these quantities.*

Definition 2. *An event in $\mathfrak{L}(1, \cdots, n)$ is said to be symmetrical if and only if it is identical with every event obtained by interchanging any pair of suffixes (i, j) $(i, j = 1, \cdots, n)$ in the definition of it.* The subset of symmetrical events in $\mathfrak{L}(1, \cdots, n)$ will be denoted by $\mathfrak{S}(1, \cdots, n)$.

From the normal form[2] of every event in $\mathfrak{L}(1, \cdots, n)$ and the principle of

---

[1] "On the probability of the occurrence of at least m events among n arbitrary events," *Annals of Math. Stat.*, Vol. 12, 1941.

[2] See Hilbert-Ackermann, *Grundzüge der theoretischen Logik*, Chap. 1.

total probabilities, we can easily see the truth of the following theorems, which may of course be made more precise.

THEOREM. *The system of* $p_{[(\alpha)]}$, $(\alpha) \in (\nu)$, 2^n *in number, forms a fundamental system for* $\mathfrak{L}(1, \cdots, n)$.

THEOREM. *The system of* $p_{[a]}((\nu))$, $0 \leqq a \leqq n$, $n+1$ *in number, forms a fundamental system for* $\mathfrak{S}(1, \cdots, n)$.

Next, a theorem of Broderick[3], in a less precise form, may be stated:

The system of $p_{((\alpha))}$ $(p_{((0))} = 1)$, $(\alpha) \in (\nu)$, 2^n *in number, forms a fundamental system for* $\mathfrak{L}$.

We may add in an easy way the following

THEOREM. *The system of* $S_a((\nu))$ $S_0((\nu)) = 1$, $0 \leqq a \leqq n$, $n+1$ *in number, forms a fundamental system for* $\mathfrak{S}$.

In the present paper we shall prove, *inter alia*, the following four theorems of the above type, stated in more precise forms.

THEOREM 1. *For any* E *in* $\mathfrak{L}$, *we have*

$$P(E) = c_0 + \sum_{\substack{(\alpha) \in (\nu) \\ a \neq 0}} c_a p_1((\alpha)),$$

where $c_0 = 0$ *or* 1 *and the* c_a*'s are integers; and they are unique*[4].

THEOREM 2. *For any* E *in* $\mathfrak{S}$, *we have*

$$P(E) = c_0 + \sum_{a=1}^{n} c_a P_a^{(1)},$$

where $c_0 = 0$ *or* 1 *and the* c_a*'s are integers; and they are unique.*

THEOREM 3. *For any* E *in* $\mathfrak{L}$, *we have*

$$P(E) = d_0 + \sum_{\substack{(\alpha) \in (\nu) \\ a \neq 0}} d_a p_{[1]}((\alpha)),$$

where $d_0 = 0$ *or* 1 *and the* d_a*'s are rational numbers and they are unique.*

THEOREM 4. *For any* E *in* $\mathfrak{S}$, *we have*

$$P(E) = d_0 + \sum_{a=1}^{n} d_a p_a^{[1]},$$

where $d_0 = 0$ *or* 1 *and the* d_a*'s are rational numbers; and they are unique.*

Less precisely, we may say that the system of $p_1((\alpha))$ or $p_{[1]}((\alpha))$ forms a fundamental system for $\mathfrak{L}$; the system of $P_a^{(1)}((\nu))$ or $P_a^{[1]}((\alpha))$ forms a fundamental system for $\mathfrak{S}$.

In fact however, we shall give much more than the mere proofs of

---

[3] Fréchet, "Compléments à un théorème de T. S. Broderick concernant les événements dependants," *Proc. Edinburgh Math. Soc.*, Ser. 2, Vol. 6 (1939).

[4] "Unique" in the sense that it is impossible to replace therein the coefficients c by other numbers which are independent of the Borel set of events and the probability function.

these theorems. We shall establish the following explicit formulas for the general parameter m.

$$\text{(i)} \qquad p_{[(0)]} = 1 - p_1((\nu)),$$

$$\text{(1.1)} \quad \text{(ii)} \qquad p_{[(\alpha)]} = \sum_{\substack{(\beta)\,\epsilon\,(\alpha) \\ n-a+b\neq 0}} (-1)^{b-1} p_1((\nu) - (\alpha) + (\beta)),^5 \qquad 1 \leq a \leq n.$$

$$p_{[(\alpha)]} = (-1)^m \frac{m-1}{n-1} \sum_{c=m}^{n} \sum_{d=\max(0,c-a)}^{\min(c,n-a)} (-1)^{c-d} \binom{n-2}{a+d-m}^{-1}$$

$$\text{(1)} \qquad \sum_{\substack{(\delta)\,\epsilon\,(\nu)-(\alpha) \\ (\gamma)-(\delta)\,\epsilon\,(\alpha)}} p_m((\gamma) - (\delta) + (\delta)), \qquad n \geq a \geq m \geq 2.^5$$

$$\text{(2.1)} \qquad p_{[a]}((\nu)) = \sum_{\substack{b=n-a \\ b\neq 0}}^{n} (-1)^{b-n+a} \binom{b}{n-a} P_b^{(1)}((\nu)), \qquad 1 \leq a \leq n.$$

$$\text{(2)} \qquad p_{[a]}((\nu)) = \sum_{b=m}^{n} (-1)^{b-m} L(n,a,b,m) P_b^{(m)}((\nu)), \qquad n \geq a \geq m \geq 2,$$

where

$$L(n,a,b,m) = \begin{cases} 0, & b < n-a+m-1, \\ (-1)^{n-a}\binom{a}{m-1}^{-1}, & b = n-a+m-1, \\ \dfrac{(-1)^{n-a}(m-1)!\,(b-m)!\cdot(a-m)!\,\{ab-n(m-1)\}}{a!\,(n-a)!\,(a+b-n-m+1)!}, & b > n-a+m-1. \end{cases}$$

$$\text{(3)} \quad \text{(i)} \qquad p_{[(0)]}((\nu)) = 1 - \frac{1}{n}\sum_{c=1}^{n} \binom{n-1}{c-1}^{-1} P_c^{[1]}.$$

$$\text{(ii)} \qquad p_{[(\alpha)]} = (-1)^m \frac{m}{n} \sum_{c=m}^{n} \sum_{d=\max(0,c-a)}^{\min(c,n-a)} (-1)^{c-d} \binom{n-1}{a+d-m}^{-1}$$

$$\sum_{\substack{(\delta)\,\epsilon\,(\nu)-(\alpha) \\ (\gamma)-(\delta)\,\epsilon\,(\alpha)}} p_{[m]}((\gamma) - (\delta) + (\delta)), \qquad n \geq a \geq m \geq 1.$$

$$\text{(4)} \qquad p_{[a]}((\nu)) = \sum_{b=m+n-a}^{n} (-1)^{n-a+b-m} \binom{b-m}{n-a}\binom{a}{m}^{-1} P_b^{[m]}((\nu)), \quad n \geq a \geq m \geq 1.$$

A simpler derivation of (1) than that given in an earlier paper[1] follows. Let us write Poincaré's formula as follows:

$$p_m((\beta)) = \sum_{c=m}^{b} (-1)^{c-m} \binom{c-1}{m-1} S_c((\beta)).$$

---

[5] Obviously we mean $((\nu) - (\alpha)) + (\beta)$ and $((\gamma) - (\delta)) + (\delta)$ respectively; similarly in the sequel.

Then for a fixed $b \geqq m$, summing over all $(\beta) \in (\nu)$, we get

$$\sum_{(\beta)\in(\nu)} p_m((\beta)) = \sum_{c=m}^{b} (-1)^{c-m} \binom{c-1}{m-1}\binom{n-c}{b-c} S_c((\nu)).$$

Hence

$$\begin{aligned}\sum_{b=m}^{n} (-1)^{b-m} \sum_{(\beta)\in(\nu)} p_m((\beta)) &= \sum_{c=m}^{n} \binom{c-1}{m-1} S_c((\nu)) \sum_{b=c}^{n} (-1)^{b-c} \binom{n-c}{b-c} \\ &= \sum_{c=m}^{n} \binom{c-1}{m-1} S_c((\nu)) \begin{cases} 1 & \text{if } c = n \\ 0 & \text{if } c < n \end{cases} \\ &= \binom{n-1}{m-1} S_n((\nu)) = \binom{n-1}{m-1} p((\nu)).\end{aligned} \tag{1}$$

A change of notation gives, for $a + b \geqq m$,

$$\binom{a+b-1}{m-1} p_{((\alpha)+(\beta))} = \sum_{c=m}^{a+b} (-1)^{c-m} \sum_{(\gamma)\in(\alpha)+(\beta)} p_m((\gamma)).$$

Hence

$$\begin{aligned}&\binom{a+b-1}{m-1} \sum_{(\beta)\in(\nu)-(\alpha)} p_{((\alpha)+(\beta))} \\ &\quad = \sum_{c=m}^{a+b} (-1)^{c-m} \sum_{d=\max(0,c-a)}^{\min(c,n-a)} \binom{n-a-d}{b-d} \sum_{\substack{(\delta)\in(\nu)-(\alpha) \\ (\gamma)-(\delta)\in(\alpha)}} p_m((\gamma) - (\delta) + (\delta)).\end{aligned}$$

Substituting in the well-known formula, for $a \geqq 1$

$$p_{[(\alpha)]} = \sum_{b=0}^{n-a} (-1)^b \sum_{(\beta)\in(\nu)-(\alpha)} p_{((\alpha)+(\beta))},$$

we get for $n \geqq a \geqq m$

$$\begin{aligned}p_{[(\alpha)]} = &\sum_{c=m}^{n} (-1)^{c-m} \sum_{d=\max(0,c-a)}^{\min(c,n-a)} \\ &\sum_{\substack{(\delta)\in(\nu)-(\alpha) \\ (\gamma)-(\delta)\in(\alpha)}} p_m((\gamma) - (\delta) + (\delta)) \left\{ \sum_{b=0}^{n-a} (-1)^b \binom{n-a-d}{b-d} \binom{a+b-1}{m-1}^{-1} \right\}.\end{aligned} \tag{1}$$

Thus the problem reduces to the summation of the following series:

$$\sum_{b=0}^{n-a} (-1)^b \binom{n-a-d}{b-d} \binom{a+b-1}{m-1}^{-1}.$$

Case 1: $m = 1$. In this case the series reduces to

$$\sum_{b=0}^{n-a} (-1)^b \binom{n-a-d}{b-d} = \begin{cases} (-1)^{n-a} & \text{if } d = n-a, \\ 0 & \text{if } d < n-a. \end{cases}$$

Hence for $a \geqq 1$,

$$p_{[(a)]} = \sum_{c=\max(1,n-a)}^{n} (-1)^{c-1} \sum_{(\gamma)-((\nu)-(\alpha))\,\epsilon\,(\alpha)} p_1((\nu)-(\alpha)+(\gamma)-((\nu)-(\alpha)))(-1)^{n-a}$$

Writing $(\gamma) - ((\nu) - (\alpha)) = (\beta)$, we obtain

$$p_{[(a)]} = \sum_{b=\max(1-n+a,0)}^{a} (-1)^{b-1} \sum_{(\beta)\,\epsilon\,(\alpha)} p_1((\nu) - (\alpha) + (\beta)).$$

This is equivalent to (1.1), (ii), while (i) is trivial.

Case 2: $m \geqq 2$. We have, for $c \geqq 1$,

$$\sum_{l=0}^{a} (-1)^l \binom{a}{l} \binom{b+l}{c}^{-1} = \frac{c}{a+b} \binom{a+b-1}{b-c}^{-1},$$

which is easily proved by induction on a.

Hence for $m \geqq 2$,

$$\begin{aligned}
&\sum_{b=0}^{n-a} (-1)^b \binom{n-a-d}{b-d} \binom{a+b-1}{m-1}^{-1} \\
&\qquad = \sum_{b'=-d}^{n-a-d} (-1)^{d+b'} \binom{n+a-d}{b'} \binom{a+b'+d-1}{m-1}^{-1} \\
&\qquad = (-1)^d \sum_{b'=0}^{n-a-d} (-1)^{b'} \binom{n-a-d}{b'} \binom{a+d-1+b'}{m-1}^{-1} \\
&\qquad = (-1)^d \frac{m-1}{n-1} \binom{n+2}{a+d-m}^{-1}
\end{aligned}$$

Substituting in (1) we get formula (1).

To derive formula (2.1) for a fixed a, $1 \leqq a \leqq n$, we sum (1.1, ii), which gives

$$p_{[a]}((\nu)) = \sum_{(\alpha)\,\epsilon\,(\nu)} p_{[(\alpha)]} = \sum_{\substack{b=0 \\ n-a+b\neq 0}}^{a} (-1)^{b-1} \sum_{(\alpha)\,\epsilon\,(\nu)} \sum_{(\beta)\,\epsilon\,(\alpha)} p_1((\nu) - (\alpha) + (\beta)).$$

Letting $(\nu) - (\alpha) + (\beta) = (\gamma)$, we get

$$p_{[a]}((\nu)) = \sum_{c=\max(1,n-a)}^{n} (-1)^{n-a+c-1} \binom{c}{n-a} \sum_{(\gamma)\,\epsilon\,(\nu)} p_1((\gamma)),$$

which is formula (2.1).

The following form of Poincaré's formula is of assistance in deriving (2):

$$p_{[a]}((\nu)) = \sum_{c=a}^{n} (-1)^{c-a} \binom{c}{a} S_a((\nu)).$$

Substituting from (1), we get

$$p_{[a]}((\nu)) = \sum_{c=a}^{n} (-1)^{c-a}\binom{c}{a}\binom{c-1}{m-1}^{-1} \sum_{b=m}^{c} (-1)^{b-m}\binom{n-b}{c-b} P_b^{(m)}((\nu))$$

$$= \sum_{b=m}^{n} (-1)^{b-m} P_b^{(m)}((\nu)) \left\{ \sum_{c=\max(a,b)}^{n} (-1)^{c-a}\binom{c}{a}\binom{n-b}{c-b}\binom{c-1}{m-1}^{-1} \right\}.$$

Thus the problem reduces to the summation of the following series:

$$L(n, a, b, m) = \sum_{c=\max(a,b)}^{n} (-1)^{c-a}\binom{c}{a}\binom{n-b}{c-b}\binom{c-1}{m-1}^{-1}$$

First, we have, for $z \geqq 0$, $y \geqq w$,

$$\sum_{x=\max(0,1-w)}^{z} (-1)^x\binom{z}{x}(x+y)\cdots(x+w)$$

$$= \begin{cases} 0 & \text{if } y-w+1 < z, \\ \dfrac{(-1)^z y!(y+1-w)!}{(z+w-1)!(y+1-w-z)!} & \text{if } y-w+1 \geqq z, \end{cases}$$

which may be easily proved by induction on z.
Next, we have

$$L(n, a, b, m) = \frac{(m-1)!}{a!} \sum_{c=\max(a,b)}^{n} (-1)^{c-a}\binom{n-b}{c-b}\frac{c(c-m)!}{(c-a)!}$$

$$= \frac{(m-1)!}{a!} \sum_{c'=\max(0,a-b)}^{n-b} (-1)^{c'+b-a}\binom{n-b}{c'}\frac{(c'+b)(c'+b-m)!}{(c'+b-a)!}$$

$$= (-1)^{b-a}\frac{(m-1)!}{a!} \sum_{c'=\max(0,a-b)}^{n-b} (-1)^{c'} \cdot \binom{n-b}{c'}\frac{(c'+b-m+1)! + (m-1)(c'+b-m)!}{(c'+b-a)!}$$

$$= (-1)^{b-a}\frac{(m-1)!}{a!}\{T(n, a, b, m) + (m-1)T(n, a, b, m+1)\},$$

where

$$T(n, a, b, m) = \sum_{c=\max(0,a-b)}^{n-b} (-1)^c\binom{n-b}{c}\frac{(c+b-m+1)!}{(c+b-a)!}$$

$$= \begin{cases} 0 & \text{if } b < n-a+m-1, \\ \dfrac{(-1)^{n-b}(a-m+1)!(b-m+1)!}{(n-a)!(a+b-n-m+1)!} & \text{if } b \geqq n-a+m-1, \end{cases}$$

by the preceding formula. Thus we get the explicit expression for $L(n, a, b, m)$ given in formula (2), which is thereby proved.

The derivations of formulas 3 and 4 are similar to the above and may be omitted.

Now we can give the essential argument for Theorems 1–4. It is evident that for any E in $\mathfrak{L}$, we have

$$P(E) = \Sigma p_{[(\alpha)]},$$

where the summation extends to certain combinations $(\alpha) \in (\nu)$. Substituting from formula (1.1) we get Theorem 1; substituting from formula (3) we get Theorem 3. Next, for any E in $\mathfrak{S}$, we have

$$P(E) = \Sigma p_{[a]}((\nu)),$$

where the summation extends to certain values of a. Substituting from formula (1.1), (i) and formula (2) we get Theorem 2; substituting from formula (2), (i) and formula (4) we get Theorem 4. We may note these proofs are "constructive".

It remains to prove the uniqueness of the coefficients in Theorems 1–4. For Broderick's theorem this has been done by Fréchet[3], by introducing "independent events". Our proof will be based on the conditions of existence, also initiated by Fréchet[6], for the systems $p_1((\alpha))$, $p_{[1]}((\alpha))$, $P_a^{(1)}((\nu))$, $P_a^{[1]}((\nu))$.

The conditions of existence of the system $p_1((\alpha))$ have been given by the author in the paper[1], though the proof there is not quite complete.

1. *Conditions of existence of the system* $P_a^{(1)}((\nu))$. Given n quantities $Q_a^{(1)}$, $1 \leqq a \leqq n$; what are the necessary and sufficient conditions that they may be the system of $P_a^{(1)}((\nu))$'s, $1 \leqq a \leqq n$, of a probability function defined over $\mathfrak{S}(1, \cdots, n)$?

From formula (1.1), (i) and formula (2) it is evident that necessary conditions are, for $1 \leqq a \leqq n$,

$$\sum_{\substack{b=n-a \\ b\neq 0}}^{n} (-1)^{b-n+a-1} \binom{b}{n-a} Q_b^{(1)} \geqq 0, \tag{3}$$

$$1 - Q_n^{(1)} \geqq 0,$$

and

$$\sum_{a=1}^{n} \sum_{\substack{b=n-a \\ b\neq 0}}^{n} (-1)^{b-n+a-1} \binom{b}{n-a} Q_b^{(1)} + 1 - Q_n^{(1)} = 1. \tag{4}$$

The last condition can be re-written as

$$\sum_{b=1}^{n} (-1)^{b-1} Q_b^{(1)} \sum_{a=\max(1,n-b)}^{n} (-1)^{n-a} \binom{b}{n-a} + 1 - Q_n^{(1)} = 1,$$

which reduces to the identity $1 = 1$.

---

[6] "Conditions d'existence de système d'événements associés à certaines probabilités," *Jour. de Math.*, 1940. However, our interpretation of the term would mean instead "conditions of existence of a probability function defined over a Borel set of events, etc."

To show that the conditions (3) are sufficient, put

$$p_{[a]} = \sum_{b=n-a}^{n} (-1)^{b-n+a-1} \binom{b}{n-a} Q_b^{(1)},$$

$$p_{[0]} = 1 - Q_n^{(1)}.$$

By (3) and (4) we have, for $0 \leqq a \leqq n$,

$$p_{[a]} \geqq 0 \quad \text{and} \quad \sum_{a=0}^{n} p_{[a]} = 1.$$

Hence they are actually the $p_{[a]}((\nu))$'s of a probability function. We want to show that the $P_a^{(1)}((\nu))$'s of this probability function coincide with the given $Q_a^{(1)}$'s, so that this is the probability function we seek. We have,

$$P_b^{(1)}((\nu)) = \sum_{(\beta)\,\epsilon\,(\nu)} p_1((\beta)) = \sum_{a=1}^{n} p_{[a]} \sum_{h=\max(1,b-n+a)}^{\min(a,b)} \binom{a}{h}\binom{n-a}{b-h}$$

$$= \sum_{c=0}^{n} \left\{ \sum_{a=\max(1,n-c)}^{n} (-1)^{c-n+a-1} \binom{c}{n-a} \sum_{h=\max(1,b-n+a)}^{\min(a,b)} \binom{a}{h}\binom{n-a}{b-h} \right\} Q_c^{(1)}.$$

Now the series in curl brackets

$$= \sum_{a=\max(1,n-c)}^{n-b} (-1)^{c-n+a-1} \binom{c}{a} \left\{ \binom{n}{b} - \binom{n-a}{b} \right\}$$

$$+ \sum_{a=n-b+1}^{n} (-1)^{c-n+a-1} \binom{c}{n-a}\binom{n}{b}$$

$$= \sum_{a=\max(1,n-c)}^{n} (-1)^{c-n+a-1} \binom{c}{n-a}\binom{n}{b}$$

$$- \sum_{a=\max(1,n-c)}^{n-b} (-1)^{c-n+a-1} \binom{c}{n-a}\binom{n-a}{b}.$$

If $c = n$, the last

$$= \binom{n}{b} - \sum_{a=1}^{n-b} (-1)^{a-1} \binom{n}{n-b}\binom{n-b}{a}$$

$$= \binom{n}{b} - \binom{n}{b} \sum_{a=1}^{n-b} (-1)^{a-1} \binom{n-b}{a} = \begin{cases} 1 & \text{if } b = n; \\ 0 & \text{if } b \neq n. \end{cases}$$

If $c < n$, we have

$$= 0 + (-1)^c \sum_{a=n-c}^{n-b} (-1)^{n-a} \binom{c}{n-a}\binom{n-a}{b}$$

$$= (-1)^c \sum_{a'=c}^{b} (-1)^{a'} \binom{c}{a'}\binom{a'}{b} = \begin{cases} 1 & \text{if } b = c; \\ 0 & \text{if } b \neq c. \end{cases}$$

Therefore

$$P_b^{(1)}((\nu)) = Q_b^{(1)}.$$

2. *Conditions of existence of the system* $p_{[1]}((\alpha))$. Given $2^n - 1$ quantities $q_{[1]}((\alpha))$, $(\alpha) \epsilon (\nu)$, $a \geqq 1$, what are the necessary and sufficient conditions that they may be the system of $p_{[1]}((\alpha))$'s, of a probability function defined over $\mathfrak{L}(1, \cdots, n)$?

From formula 3 it is evident that necessary conditions are

$$\frac{1}{n}\sum_{c=1}^{n}\sum_{d=\max(0,c-a)}^{\min(c,n-a)}(-1)^{c-d-1}\binom{n-1}{a+d-1}^{-1}\sum_{\substack{(\delta)\,\epsilon\,(\nu)-(\alpha)\\(\gamma)-(\delta)\,\epsilon\,(\alpha)}} q_{[1]}((\gamma)-(\delta)+(\delta)) \geqq 0, \tag{5}$$

$$1-\frac{1}{n}\sum_{c=1}^{n}\binom{n-1}{c-1}^{-1}\sum_{(\gamma)\,\epsilon\,(\nu)} p_{[1]}((\gamma)) \geqq 0;$$

and

$$1+\frac{1}{n}\sum_{(\alpha)\,\epsilon\,(\nu)}\sum_{c=1}^{n}\sum_{d=\max(0,c-a)}^{\min(c,n-a)}(-1)^{c-d-1}\binom{n-1}{a+d-1}^{-1}\sum_{\substack{(\delta)\,\epsilon\,(\nu)-(\alpha)\\(\gamma)-(\delta)\,\epsilon\,(\alpha)}} q_{[1]}((\gamma)-(\delta)+(\delta)) = 1. \tag{6}$$

Consider the sum

$$\sum_{(\alpha)\,\epsilon\,(\nu)}\sum_{d=\max(0,c-a)}^{\min(c,n-a)}(-1)^{d}\binom{n-1}{a+d-1}^{-1}\sum_{\substack{(\delta)\,\epsilon\,(\nu)-(\alpha)\\(\gamma)-(\delta)\,\epsilon\,(\alpha)}} q_{[1]}((\gamma)-(\delta)+(\delta)).$$

For a fixed (δ), the number of ways of writing $(\gamma) = (\gamma) - (\delta) + (\delta)$ is $\binom{c}{d}$, then since $(\gamma) - (\delta) \epsilon (\alpha)$ but $(\alpha) - ((\gamma) - (\delta)) \epsilon (\nu) - (\gamma)$, the number of choices of (α) is $\binom{n-c}{a-c+d}$. Thus the coefficient of $q_{[1]}((\gamma))$ in the sum is

$$\sum_{a=0}^{n}\sum_{d=\max(0,c-a)}^{\min(c,n-a)}(-1)^{d}\binom{c}{d}\binom{n-c}{a-c+d}\binom{n-1}{a+d-1}^{-1} = \binom{n-1}{c-1}^{-1}\sum_{a=0}^{n}\sum_{d=\max(0,c-a)}^{\min(c,n-a)}(-1)^{d}\binom{c}{d}\binom{a+d-1}{c-1} = 0.$$

Therefore the condition (6) reduces to the identity $1 = 1$.

To show that conditions (6) are sufficient, put the left-hand sides of (5) equal to $p_{[(a)]}$ and $p_{[(0)]}$ respectively. Then

$$\begin{aligned} p_{[(a)]} &= \sum_{(\beta)\,\epsilon\,(\nu)-(\alpha)} p_{[(\alpha)+(\beta)]} \\ &= \frac{1}{n}\sum_{c=1}^{n}(-1)^{c-1}\sum_{b=0}^{n-a}\sum_{d=\max(0,c-a-b)}^{\min(c,n-a-b)}(-1)^{d}\binom{n-1}{a+d-1}^{-1}\sum_{(\beta)\,\epsilon\,(\nu)-(\alpha)}\sum_{\substack{(\delta)\,\epsilon\,(\nu)-(\alpha)-(\beta)\\(\gamma)-(\delta)\,\epsilon\,(\alpha)+(\beta)}} q_{[1]}((\gamma)-(\delta)+(\delta)). \end{aligned} \tag{7}$$

Let $(\gamma) = (\gamma) - (\phi) + (\phi)$, where $(\phi) \in (\alpha)$, $(\gamma) - (\phi) \in (\nu) - (\alpha)$. Then the sum in the curl brackets can be written, by a combinatorial calculation, as

$$\sum_{f=0}^{\min(a,c)} \left\{ \sum_{b=0}^{n-a} \sum_{d=\max(0,c-f-b)}^{\min(c-f,n-a-b)} (-1)^d \binom{c-f}{d} \binom{n-a-c+f}{b-c+d+f} \binom{n-1}{a+b+d-1}^{-1} \right\} \sum_{\substack{(\phi)\in(\alpha) \\ (\gamma)-(\phi)\in(\nu)-(\alpha)}} q_{[1]}((\gamma) - (\phi) + (\phi)).$$

The sum in the last curl brackets is

$$\binom{n-1}{a+c-f-1}^{-1} \sum_{b=0}^{n-a} \sum_{d=\max(0,c-f-b)}^{\min(c-f,n-a-b)} (-1)^d \binom{c-f}{d} \binom{a+b+d-1}{a+c-f-1}.$$

Inverting the order of summations,

$$\binom{n-1}{a+c-f-1}^{-1} \sum_{d=\max(0,c-f-n+a)}^{\min(c-f,n-a)} (-1)^d \binom{c-f}{d} \sum_{b=c-f-d}^{n-a-d} \binom{a+b+d-1}{a+c-f-1}$$

$$= \binom{n-1}{a+c-f-1}^{-1} \sum_{d=\max(0,c-f-n+a)}^{\min(c-f,n-a)} (-1)^d \binom{c-f}{d} \binom{n}{a+c-f}$$

$$= \binom{n}{a+c-f} \binom{n-1}{a+c-f-1}^{-1} \sum_{d=0}^{c-f} (-1)^d \binom{c-f}{d} = \begin{cases} \frac{n}{a} & \text{if } f = c, \\ 0 & \text{if } f \neq c. \end{cases}$$

Hence (7) reduces to

$$p_{((a))} = \frac{1}{a} \sum_{c=1}^{n} (-1)^{c-1} \sum_{(\gamma)\in(\alpha)} q_{[1]}((\gamma)).$$

Then

$$S_b((\alpha)) = \sum_{(\beta)\in(\alpha)} p_{((\beta))} = \frac{1}{b} \sum_{\substack{(\delta)\in(\alpha) \\ d\neq 0}} (-1)^{d-1} \binom{a-d}{b-d} q_{[1]}((\delta))$$

$$p_{[1]}((\alpha)) = \sum_{b=1}^{a} (-1)^{b-1} S_b((\alpha))$$

$$= \sum_{\substack{(\delta)\in(\alpha) \\ d\neq 0}} \left\{ \sum_{b=1}^{a} (-1)^{b-d} \binom{a-d}{b-d} \right\} q_{[1]}((\delta)) = q_{[1]}((\alpha)).$$

The conditions of existence of the system $P_a^{[1]}((\nu))$, $1 \leqq a \leqq n$, are similarly deduced from formula (3), (i) and formula (4) with $m = 1$.

Now we can prove the uniqueness of the coefficients in Theorems 1–4. Since the proofs are all exactly similar, we take Theorem 2. Suppose, if possible, there exists another system of coefficients c_a', $0 \leqq a \leqq n$ so that

$$P(E) = c_0 + \sum_{a=1}^{n} c_a P_a^{(1)}((\nu)) = c_0' + \sum_{a=1}^{n} c_a' P_a^{(1)}((\nu)).$$

Taking the difference, we get a linear polynomial in the variables $P_a^{(1)}((\nu))$, $1 \leqq a \leqq n$ which must vanish:

$$(8) \qquad (c_0 - c_0') + \sum_{a=1}^{n} (c_a - c_a')P_a^{(1)}((\nu)) = 0,$$

for all "admissible" values of the variables. These values, say $Q_a^{(1)}$, are precisely those which satisfy the conditions (3).

It is evidently easy to construct a system of $Q_a^{(1)}$, $1 \leqq a \leqq n$, which satisfy the conditions (3) written with the sign of strict inequality ">". Hence in a sufficiently small neighborhood of the point $(Q_1^{(1)}, Q_2^{(1)}, \cdots, Q_n^{(1)})$ in the n-dimensional space these strict inequalities still hold. Hence the polynomial vanishes in this neighborhood and so must vanish identically; that is,

$$c_a - c_a' = 0 \quad \text{for} \quad 0 \leqq a \leqq n. \qquad \text{Q. E. D.}$$

ON THE EFFICIENT DESIGN OF STATISTICAL INVESTIGATIONS

By Abraham Wald

Columbia University

1. Introduction. A theory of efficient design of statistical investigations has been developed by R. A. Fisher[1] and his followers mainly in connection with agricultural experimentation. However, the same methods can be applied to other fields also. All statistical designs treated in the aforementioned theory refer to problems of testing linear hypotheses. By testing a linear hypothesis we mean the following problem: Let $y_1, \cdots, y_N$ be N independently and normally distributed variates with a common variance σ^2. It is assumed that the expected value of y_α is given by

$$E(y_\alpha) = \beta_1 x_{1\alpha} + \beta_2 x_{2\alpha} + \cdots + \beta_p x_{p\alpha} \qquad (\alpha = 1, \cdots, N) \tag{1}$$

where the quantities $x_{i\alpha} (i = 1, \cdots, p; \alpha = 1, \cdots, N)$ are known constants and $\beta_1, \cdots, \beta_p$ are unknown constants. The coefficients $\beta_1, \cdots, \beta_p$ are called the population regression coefficients of y on $x_1, x_2, \cdots$, and x_p, respectively. The hypothesis that the unknown regression coefficients $\beta_1, \cdots, \beta_p$ satisfy a set of linear equations

$$g_{i1}\beta_1 + \cdots + g_{ip}\beta_p = g_i \qquad (i = 1, \cdots, r; r \leqq p), \tag{2}$$

is called a linear hypothesis. The problem under consideration is that of testing the hypothesis (2) on the basis of the observed values $y_1, \cdots, y_N$.

In many cases the experimenter has a certain amount of freedom in the choice of the values $x_{i\alpha}$. The efficiency of the test is greatly affected by the values of $x_{i\alpha}$. The statistical investigation is efficiently designed if the values $x_{i\alpha}$ are chosen so that the sensitivity of the test is maximized. Let us illustrate this by a simple example. Suppose that x and y have a bivariate normal distribution and we want to test the hypothesis that the regression coefficient β of y on x has a particular value β_0. Suppose, furthermore, that the test has to be carried out on the basis of N pairs of observations $(x_1, y_1), \cdots, (x_N, y_N)$, where the experiments are performed in such a way that $x_1, \cdots, x_N$ are not random variables but have predetermined fixed values. It is known that the variance of the least square estimate b of β is inversely proportional to $\sum_{\alpha=1}^{N} (x_\alpha - \bar{x})^2$ where $\bar{x} = (x_1 + \cdots + x_N)/N$. Hence, if we can freely choose the values $x_1, \cdots, x_N$ in a certain domain D, the greatest sensitivity of the test will be achieved by choosing $x_1, \cdots, x_N$ so that $\Sigma(x_\alpha - \bar{x})^2$ becomes a maximum.

In the next section we will introduce a measure of the efficiency of the design

[1] See for instance R. A. Fisher, *The Design of Experiments*, Oliver and Boyd, London, 1935.

of a statistical investigation for testing a linear hypothesis. In sections 3 and 4 it will be shown that some well known experimental designs, used widely in agricultural experimentation, are most efficient in the sense of the definition given in section 2.

2. A measure of the efficiency of the design of a statistical investigation for testing a linear hypothesis. The hypothesis (2) can be reduced by a suitable linear transformation to the canonical form

$$\beta_1 = \beta_2 = \cdots = \beta_r = 0, \qquad (r \leq p). \tag{3}$$

Hence, we can restrict ourselves without loss of generality to the consideration of the hypothesis (3).

Denote $\sum_{\alpha=1}^{N} x_{i\alpha}x_{j\alpha}$ by a_{ij} and let the matrix $\| c_{ij} \|$ be the inverse of the matrix $\| a_{ij} \|$ $(i, j = 1, \cdots, p)$. Denote by b_i the least square estimate of β_i $(i = 1, \cdots, p)$. It is known that the estimates $b_1, \cdots, b_p$ have a joint normal distribution with mean values $\beta_1, \cdots, \beta_p$, respectively. It is furthermore known that the covariance of b_i and b_j is equal to $c_{ij}\sigma^2$. The statistic used for testing the hypothesis (3) is given by

$$F = \frac{N - p}{r} \frac{\sum_{l=1}^{r}\sum_{m=1}^{r} a^*_{lm} b_l b_m}{\sum_{\alpha=1}^{N} (y_\alpha - b_1 x_{1\alpha} \cdots - b_p x_{p\alpha})^2} \tag{4}$$

where $\| a^*_{lm} \|$ is the inverse of $\| c_{lm} \|$ $((l, m = 1, \cdots, r)$. The statistic F has the F-distribution with r and $N - p$ degrees of freedom. The critical region for testing the hypothesis (3) is given by the inequality

$$F \geq F_0, \tag{5}$$

where the constant F_0 is determined so that the probability that $F \geq F_0$ (calculated under the assumption that (3) holds) is equal to the level of significance we wish to have.

It is known that the power function[2] of the critical region (5) depends only on the single parameter

$$\lambda = \frac{1}{\sigma^2} \sum_{l=1}^{r}\sum_{m=1}^{r} a^*_{lm}\beta_l\beta_m. \tag{6}$$

Furthermore this power function is a monotonically increasing function of λ. The coefficients a^*_{lm} are functions of the quantities $x_{i\alpha}$ $(i = 1, \cdots, p; \alpha = 1, \cdots, N)$. The choice of the values $x_{i\alpha}$ $(i = 1, \cdots, p; \alpha = 1, \cdots, N)$ is the better the greater the corresponding value of λ. If $r = 1$, the expression λ

[2] See for instance P. C. TANG, "The power function of the analysis of variance tests," *Stat. Res. Mem.*, Vol. II, 1938.

reduces to $\frac{1}{\sigma^2} a_{11}^* \beta_1^2$. Hence, if $r = 1$, we maximize λ by maximizing a_{11}^*. Since $a_{11}^* = 1/c_{11}$, we maximize λ by minimizing c_{11}. Thus, if $r = 1$, we can say that we obtain the most powerful test by minimizing c_{11}, i.e. by minimizing the variance of b_1. If $r > 1$, the difficulty arises that no set of values $x_{i\alpha}$ $(i = 1, \cdots, p; \alpha = 1, \cdots, N)$ can be found for which λ becomes a maximum irrespective of the values of the unknown parameters $\beta_1, \cdots, \beta_r$. Hence, if $r > 1$, we have to be satisfied with some compromise solution. For this purpose let us consider the unit sphere

$$\beta_1^2 + \cdots + \beta_r^2 = 1, \tag{7}$$

in the space of the parameters $\beta_1, \cdots, \beta_r$. It is known that the smallest root in ρ of the determinantal equation

$$\begin{vmatrix} a_{11}^* - \rho & a_{12}^* & \cdots & a_{1r}^* \\ a_{21}^* & a_{22}^* - \rho & \cdots & a_{2r}^* \\ \cdot & \cdot & \cdots & \cdot \\ a_{r1}^* & a_{r2}^* & \cdots & a_{rr}^* - \rho \end{vmatrix} = 0, \tag{8}$$

is equal to the minimum value of $\sigma^2\lambda$ on the unit sphere (7). Similarly the greatest root of (8) is equal to the maximum value of $\sigma^2\lambda$ on the sphere (7). The compromise solution of maximizing the smallest root of (8) seems to be a very reasonable one. However, for the sake of certain mathematical simplifications, we propose to maximize the product of the r roots of (8). Since the product of the roots of (8) is equal to the determinant

$$\begin{vmatrix} a_{11}^* & \cdots & a_{1r}^* \\ \cdot & \cdots & \cdot \\ a_{r1}^* & \cdots & a_{rr}^* \end{vmatrix}, \tag{9}$$

we have to maximize the determinant (9). The value of the determinant $|c_{lm}|$ $(l, m = 1, \cdots, r)$ is the reciprocal of that of (9). Hence we maximize (9) by minimizing the determinant $|c_{lm}|$. The generalized variance of the set of variates $b_1, \cdots, b_r$ is equal to the product of σ^{2r} and the determinant $|c_{lm}|$. Thus, our result can be expressed as follows: The optimum choice of the values of $x_{i\alpha}$ is that for which the generalized variance of the variates $b_1, \cdots, b_r$ becomes a minimum.

Any set of pN values $x_{i\alpha}$ $(i = 1, \cdots, p; \alpha = 1, \cdots, N)$ can be represented by a point in the pN-dimensional Cartesian space. Denote by D the set of all points in the pN-dimensional space which we are free to choose. If N is fixed and if any point of D can be equally well chosen, the following two definitions seem to be appropriate:

DEFINITION 1. *Denote by c the minimum value of the determinant $|c_{lm}|$ $(l, m = 1, \cdots, r)$ in the domain D. Then the ratio $c/|c_{lm}|$ is called the efficiency of the design of the statistical investigation for testing the hypothesis* (3).

DEFINITION 2. *The design of the statistical investigation for testing the hypothesis* (3) *is said to be most efficient if its efficiency is equal to* 1.

3. Efficiency of the Latin square design. A widely used and important design in agricultural experimentation is the so-called Latin square. Suppose we wish to find out by experimentation whether there is any significant difference among the yields of m different varieties $v_1, \cdots, v_m$. For this purpose the experimental area is subdivided into m^2 plots lying in m rows and m columns and each plot is assigned to one of the varieties $v_1, \cdots, v_m$. If each variety appears exactly once in each row and exactly once in each column, we have a Latin square arrangement. Denote by y_{ijk} the yield of the variety v_k on the plot which lies in the i-th row and j-th column. The subscript k is, of course, a single valued function of the subscripts i and j, since to each plot only one variety is assigned. The following assumptions are made: the variates y_{ijk} are independently and normally distributed with a common variance σ^2 and the expected value of y_{ijk} is given by

$$E(y_{ijk}) = \mu_i + \nu_j + \rho_k. \tag{10}$$

The parameters σ^2, μ_i, ν_j and ρ_k are unknown. The hypothesis to be tested is the hypothesis that variety has no effect on yield, i.e.

$$\rho_1 = \rho_2 = \cdots = \rho_k. \tag{11}$$

We associate the positive integer $\alpha(i, j) = (i - 1)m + j$ with the plot which lies in the i-th row and j-th column. $(i, j = 1, \cdots, m)$. It is clear that for any positive integer $\alpha \leq m^2$ there exists exactly one plot, i.e. exactly one pair of values i and j, such that $\alpha = \alpha(i, j)$. In the following discussions the symbol y_α $(\alpha = 1, \cdots, m^2)$ will denote the yield y_{ijk} where the indices i and j are determined so that $\alpha(i, j) = \alpha$. The plot in the i-th row and j-th column will be called the α-th plot where $\alpha = \alpha(i, j)$.

We define the symbols $t_{i\alpha}$, $u_{j\alpha}$, $z_{k\alpha}$ $(i, j, k = 1, \cdots, m; \alpha = 1, \cdots, m^2)$, as follows: $t_{i\alpha} = 1$ if the α-th plot lies in the i-th row, and $t_{i\alpha} = 0$ otherwise. Similarly $u_{j\alpha} = 1$ if the α-th plot lies in the j-th column, and $u_{j\alpha} = 0$ otherwise. Finally $z_{k\alpha} = 1$ if the k-th variety is assigned to the α-th plot, and $z_{k\alpha} = 0$ otherwise. Then equation (10) can be written as

$$\begin{aligned} E(y_\alpha) = {} & \mu_1 t_{1\alpha} + \cdots + \mu_m t_{m\alpha} + \nu_1 u_{1\alpha} + \cdots \\ & + \nu_m u_{m\alpha} + \rho_1 z_{1\alpha} + \cdots + \rho_m z_{m\alpha}. \end{aligned} \tag{12}$$

Denote the arithmetic means $\frac{1}{m^2}\sum_{\alpha=1}^{m^2} t_{i\alpha}$, $\frac{1}{m^2}\sum_{\alpha=1}^{m^2} u_{i\alpha}$, and $\frac{1}{m^2}\sum_{\alpha=1}^{m^2} z_{i\alpha}$ by $\bar{t}_i$, $\bar{u}_i$ and $\bar{z}_i$ respectively. Let $t'_{i\alpha} = t_{i\alpha} - \bar{t}_i$, $u'_{i\alpha} = u_{i\alpha} - \bar{u}_i$, $z'_{i\alpha} = z_{i\alpha} - \bar{z}_i$, $\mu'_i = \mu_i - \mu_m$, $\nu'_i = \nu_i - \nu_m$ and $\rho'_i = \rho_i - \rho_m$ for $i = 1, \cdots, m - 1$. Let furthermore $w_\alpha = 1$ for $\alpha = 1, \cdots, m^2$. Then we have

$$\begin{cases} t_{i\alpha} = t'_{i\alpha} + \bar{t}_i w_\alpha; \quad u_{i\alpha} = u'_{i\alpha} + \bar{u}_i w_\alpha; \quad z_{i\alpha} = z'_{i\alpha} + \bar{z}_i w_\alpha; \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (i = 1, \cdots, m - 1) \\ t_{m\alpha} = (1 - \bar{t}_1 - \cdots - \bar{t}_{m-1})w_\alpha - t'_{1\alpha} - \cdots - t'_{m-1,\alpha}, \\ u_{m\alpha} = (1 - \bar{u}_1 - \cdots - \bar{u}_{m-1})w_\alpha - u'_{1\alpha} - \cdots - u'_{m-1,\alpha}, \\ z_{m\alpha} = (1 - \bar{z}_1 - \cdots - \bar{z}_{m-1})w_\alpha - z'_{1\alpha} - \cdots - z'_{m-1,\alpha}. \end{cases} \tag{13}$$

From (12) and (13) we obtain

$$E(y_\alpha) = \xi w_\alpha + \sum_{i=1}^{m-1} \mu_i' t_{i\alpha}' + \sum_{i=1}^{m-1} \nu_i' u_{i\alpha}' + \sum_{i=1}^{m-1} \rho_i' z_{i\alpha}' \tag{14}$$

where

$$\xi = \sum_{i=1}^{m-1} \mu_i' \bar{t}_i + \sum_{i=1}^{m-1} \nu_i' \bar{u}_i + \sum_{i=1}^{m-1} \rho_i' \bar{z}_i + \mu_m + \nu_m + \rho_m .$$

The hypothesis (11) can be written as

$$\rho_1' = \rho_2' = \cdots = \rho_{m-1}' = 0. \tag{15}$$

This is a linear hypothesis in canonical form as given in (3). The values $z_{i\alpha}'$ $(i = 1, \cdots, m-1; \alpha = 1, \cdots, m^2)$ depend on the way in which the varieties $v_1, \cdots, v_m$ are assigned to the m^2 plots. We will show that we obtain a most efficient design if we distribute the varieties over the m^2 plots in a Latin square arrangement, i.e. if each variety appears exactly once in each row and exactly once in each column.

Let $q_{1\alpha} = w_\alpha$, $q_{i+1,\alpha} = t_{i\alpha}'$ $(i = 1, \cdots, m-1)$, $q_{m+j,\alpha} = u_{j\alpha}'$ $(j = 1, \cdots, m-1)$ and $q_{2m-1+k,\alpha} = z_{k\alpha}'$ $(k = 1, \cdots, m-1)$. Denote $\sum_{\alpha=1}^{m^2} q_{i\alpha} q_{j\alpha}$ by a_{ij} $(i, j = 1, 2, \cdots, 3m-2)$ and let the matrix $\| c_{ij} \|$ be the inverse of the matrix $\| a_{ij} \|$ $(i, j = 1, \cdots, 3m-2)$. Let us denote by Δ the determinant $| a_{ij} |$ $(i, j = 1, \cdots, 3m-2)$, by Δ_1 the determinant $| a_{ij} |$ $(i, j = 1, \cdots, 2m-1)$, by Δ_2 the determinant $| a_{ij} |$ $(i, j = 2m, \cdots, 3m-2)$ and Δ_2' the determinant $| c_{ij} |$ $(i, j = 2m, \cdots, 3m-2)$. We have to show that for the Latin square arrangement Δ_2' becomes a minimum. From a known theorem[3] about determinants it follows that

$$\Delta_2' = \Delta_1/\Delta. \tag{16}$$

Hence, we have merely to show that Δ/Δ_1 becomes a maximum for the Latin square arrangement. Denote by $\bar{\Delta}$, $\bar{\Delta}_1$ and $\bar{\Delta}_2$ the values taken by Δ, Δ_1 and Δ_2, respectively, in the case of a Latin square arrangement. Since, for the Latin square arrangement, as is known,

$$\sum_{\alpha=1}^{m^2} z_{k\alpha}' u_{j\alpha}' = \sum_{\alpha=1}^{m^2} z_{k\alpha}' t_{i\alpha}' = \sum_{\alpha=1}^{m^2} z_{k\alpha}' w_\alpha = 0 \qquad (i, j, k = 1, \cdots, m-1)$$

we have

$$\frac{\bar{\Delta}}{\bar{\Delta}_1} = \bar{\Delta}_2 . \tag{17}$$

Since the matrix $\| a_{ij} \|$ $(i, j = 1, \cdots, 3m-2)$ is positive definite we have

$$\frac{\Delta}{\Delta_1} \leqq \Delta_2 . \tag{18}$$

[3] See M. BÔCHER, *Introduction to Higher Algebra*, 1931, pp. 31.

Because of (17) and (18) the Latin square design is proved to be most efficient if we show that $\Delta_2 \leq \bar{\Delta}_2$.

Denote by Δ_2^* the m-rowed determinant $|a_{ij}|$ $(i, j = 1, 2m, 2m+1, \cdots, 3m-2)$. Since $a_{1j} = 0$ for $j \neq 1$, we have

$$\Delta_2^* = a_{11}\Delta_2 = m^2\Delta_2. \tag{19}$$

Denote $\sum_{\alpha=1}^{m^2} z_{i\alpha}z_{j\alpha}$ by b_{ij} $(i, j = 1, \cdots, m)$. Then

$$\begin{cases} b_{ij} = 0, & \text{for } i \neq j \\ \text{and } b_{ii} = N_i, \end{cases} \tag{20}$$

where N_i denotes the number of plots to which the variety v_i has been assigned. Because of (20) we have

$$\begin{vmatrix} b_{11} & \cdots & b_{1m} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ b_{m1} & & b_{mm} \end{vmatrix} = N_1N_2\cdots N_m. \tag{21}$$

According to (13) we have

$$\begin{aligned} z'_{i\alpha} + \bar{z}_i w_\alpha &= z_{i\alpha}, \qquad (i = 1, \cdots, m-1) \\ -z'_{1\alpha} - \cdots - z'_{m-1,\alpha} + w_\alpha(1 - \bar{z}_1 - \cdots - \bar{z}_{m-1}) &= z_{m\alpha}. \end{aligned} \tag{22}$$

The determinant of these equations is given by

$$\lambda = \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & \bar{z}_1 \\ 0 & 1 & 0 & \cdots & 0 & 0 & \bar{z}_2 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 0 & 1 & \bar{z}_{m-1} \\ -1 & -1 & -1 & \cdots & -1 & -1 & \delta \end{vmatrix} \tag{23}$$

where $\delta = 1 - \bar{z}_1 - \bar{z}_2 - \cdots - \bar{z}_{m-1}$. It is easy to verify that

$$\lambda = 1. \tag{24}$$

From (21), (22) and (24) it follows that

$$\Delta_2^* = N_1N_2\cdots N_m. \tag{25}$$

Hence, from (19) we obtain

$$\Delta_2 = N_1N_2\cdots N_m/m^2. \tag{26}$$

In the case of a Latin square design we have $N_1 = N_2 = \cdots = N_m = m$. Hence

$$\bar{\Delta}_2 = m^{m-2}. \tag{27}$$

Because of the condition $N_1 + N_2 + \cdots + N_m = m^2$, the right hand side of (26) becomes a maximum when $N_1 = N_2 = \cdots = N_m = m$. Thus $\Delta_2 \leq \bar{\Delta}_2$ and consequently the Latin square design is proved to be most efficient.

4. Efficiency of Graeco-Latin and higher squares. Consider m varieties $v_1, \cdots, v_m$ and m treatments $q_1, \cdots, q_m$ Suppose that we wish to find out by experimentation whether the yield is affected by varieties or treatments. For this purpose the experimental area is subdivided into m^2 plots lying in m rows and m columns and to each plot one of the varieties and one of the treatments is assigned. We call this arrangement a Graeco-Latin square if the following conditions are fulfilled: 1) each variety appears exactly once in each row and exactly once in each column; 2) each treatment appears exactly once in each row and exactly once in each column; 3) each variety is combined with each of the treatments exactly once.

The following general abstract scheme includes the Latin square and Graeco-Latin square as special cases: Consider an r-way classification with m classes in each classification. Denote by $y_{a_1 a_2 \cdots a_r}$ the value of a certain characteristic of an individual who is classified in the a_1-class of the first classification, in the a_2-class of the second classification, $\cdots$, and in the a_r-class of the r-th classification. Suppose that m^2 observations are made for the purpose of investigating the effect of the classes on the value of the characteristic under consideration. We will say that we have a generalized Latin square design if the following condition is fulfilled: *Let i, j, m' and m'' be an arbitrary set of four positive integers for which $i \neq j$, $i \leq r$, $j \leq r$, $m' \leq m$ and $m'' \leq m$. Then among the m^2 individuals observed there exists exactly one individual who belongs to the m'-class of the i-th classification and m''-class of the j-th classification.*

It is clear that if $r = 3$ the above scheme is a Latin square. If $r = 4$ we have a Graeco-Latin square.

Assume that the observations $y_{a_1 \cdots a_r} (a_1, a_2, \cdots, a_r = 1, \cdots, m)$ are normally and independently distributed with a common variance σ^2. Assume furthermore that the expected value of $y_{a_1 \cdots a_r}$ is given by

$$E(y_{a_1 a_2 \cdots a_r}) = \gamma_{1a_1} + \cdots + \gamma_{ra_r}.$$

The parameters σ^2 and γ_{ia} $(i = 1, \cdots, r;\ a = 1, \cdots, m)$ are unknown constants. Suppose that we wish to test the hypothesis that

$$\gamma_{i1} = \gamma_{i2} = \cdots = \gamma_{im}. \tag{28}$$

It can be shown that if the number of observations is limited to m^2, we obtain a most efficient design by constructing a generalized Latin square. The proof of this statement is similar to that of the efficiency of the Latin square and is therefore omitted.

SOME SIGNIFICANCE TESTS FOR NORMAL BIVARIATE DISTRIBUTIONS

By D. S. Villars and T. W. Anderson

United States Rubber Company, Passaic, New Jersey, and Princeton University

1. Introduction. In the theory of linear regression of y on x where y is normally distributed about a linear function of x, say $\nu + \beta x$, where x is a "fixed" variate, the t-test for the hypothesis that β is zero (that y is distributed about ν; independent of x) is well known. In this paper we apply some general statistical theory to the similar problem where x and y are jointly normally distributed. This case is commonly known as the case of "error in both variates." We derive a criterion for testing the hypothesis that the population means are the coordinates of a specified point when the ratio of the variances and the population correlation coefficient are known. When the ratio of variances is known, a criterion is derived to test whether the correlation coefficient is zero.

2. The means. Let us consider a sample of n pairs of observations $(x_1, y_1; x_2, y_2; \cdots; x_n, y_n)$ from a normal bivariate population. Let the variances of x and of y be σ_x^2 and σ_y^2, respectively; and the correlation coefficient, say ρ, be zero. Suppose the ratio of the weight of y to the weight of x, say $\gamma = w_y/w_x = \sigma_x^2/\sigma_y^2$, is known although the variances are not known. It is clear then, that $\sqrt{\gamma}\, y$ has variance σ_x^2. Since the observations y_i $(i = 1, 2, \cdots, n)$ can be transformed into revised observations $\sqrt{\gamma}\, y_i = y_i'$, we lose no generality by assuming that x and y are both distributed with variance σ^2.

Under the assumption of equality of variances and independence of variates we shall derive a criterion for testing the null hypothesis that each observation x_i is of a variate distributed about the same population mean μ and each observation y_i is of a variate distributed about the same population mean ν. The hypothesis may be stated symbolically as:

$$H_0: E(x) = \mu, \qquad E(y) = \nu,$$

given $\sigma_x^2 = \sigma_y^2 = \sigma^2$ and $\rho = 0$. We can write

$$\sum_{i=1}^{n} (x_i - \mu)^2 = n(\bar{x} - \mu)^2 + S_x,$$

$$\sum_{i=1}^{n} (y_i - \nu)^2 = n(\bar{y} - \nu)^2 + S_y,$$

where

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

$$S_x = \sum_{i=1}^{n} (x_i - \bar{x})^2, \qquad S_y = \sum_{i=1}^{n} (y_i - \bar{y})^2.$$

Then $n(\bar{x} - \mu)^2/\sigma^2$ and $n(\bar{y} - \nu)^2/\sigma^2$ are each distributed independently as χ^2 with one degree of freedom and each of S_x/σ^2 and S_y/σ^2 follow the χ^2-law with $n - 1$ degrees of freedom. If we define

$$r = \sqrt{(\bar{x} - \mu)^2 + (\bar{y} - \nu)^2}, \qquad S_r = S_x + S_y, \tag{1}$$

then nr^2/σ^2 and S_r/σ^2 have independent χ^2-distributions with 2 and $2n - 2$ degrees of freedom, respectively.

It follows from this that

$$R = \frac{nr^2}{2\sigma^2} \Big/ \frac{S_r}{(2n-2)\sigma^2} = n(n-1)\frac{r^2}{S_r} = n(n-1)\frac{(\bar{x} - \mu)^2 + (\bar{y} - \nu)^2}{S_x + S_y}, \tag{2}$$

has the F-distribution with 2 and $2n - 2$ degrees of freedom.

Let us define F_α so

$$\int_{F_\alpha}^{\infty} h_{2,2n-2}(F)\, dF = \alpha, \tag{3}$$

where $h_{2,2n-2}(F)$ is the F-distribution with 2 and $2n - 2$ degrees of freedom and $0 \leq \alpha \leq 1$. Then the probability is α that the sample statistic R is greater than or equal to F_α, i.e.,

$$P\{R \geq F_\alpha\} = \alpha. \tag{4}$$

In considering a sample value of R, at significance level α, one rejects the hypothesis of the means being μ and ν, respectively, if R is larger than F_α, i.e., larger than 1 and larger than the α significance point in Snedecor's tables [1].

This F-test is a straightforward generalization to the bivariate case of the usual t-test as applied to the univariate case. In each case the sum of squares of distances of the observations from the population mean is broken up into the sum of squares of distances from the sample mean plus n times the square of the distance from the sample mean to the population mean. The t-test for the univariate case depends on the ratio of the distance of the sample mean from the population mean to the square root of the sum of squares of distances from the observations to the sample mean. The proposed F-test depends upon the ratio of the square of the distance of the sample mean from the population mean to the sum of squares of distances from the observations to the sample mean.

It can easily be shown that the likelihood ratio criterion for this hypothesis is

$$\lambda = \left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n}(y_i - \bar{y})^2}{\sum_{i=1}^{n}(x_i - \mu)^2 + \sum_{i=1}^{n}(y_i - \nu)^2}\right]^n = \left[1 + \frac{R}{n-1}\right]^n. \tag{5}$$

The hypothesis considered here is one of a class of hypotheses treated by Kolodziejczyk [2] in a paper in which he considers the likelihood ratio criterion for a set of general linear hypotheses.

Equation (4) may be written

$$P\{(\bar{x} - \mu)^2 + (\bar{y} - \nu)^2 \geq r_\alpha^2\} = \alpha, \tag{6}$$

where $r_\alpha^2 = F_\alpha\,(S_x + S_y)/[n(n-1)]$. The probability is α that the distance from the sample means $\bar{x}, \bar{y}$ to the population means μ, ν is greater than or equal to r_α. We may call r_α the *fiducial radius* [3], and the equation $(\bar{x} - \mu)^2 + (\bar{y} - \nu)^2 = r_\alpha^2$ defines the *confidence region* for the population means.

Suppose we have two samples of n_1 and n_2 pairs of observations, respectively, from normal bivariate distributions. If the population mean of each x variate is μ and the population mean of each y variate is ν, the population variance of each variate is σ^2, and the correlation coefficient is zero, then the sample means $\bar{x}_1$ and $\bar{y}_1$ of the first sample and $\bar{x}_2$ and $\bar{y}_2$ of the second sample follow normal distributions. Also $\bar{x}_1 - \bar{x}_2$ and $\bar{y}_1 - \bar{y}_2$ are normally distributed. Then $r'^2 = n_1 n_2/(n_1 + n_2)[(\bar{x}_1 - \bar{x}_2)^2 + (\bar{y}_1 - \bar{y}_2)^2]/\sigma^2$ has the χ^2-distribution with 2 degrees of freedom. Let

$$S'_{r'} = \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2 + \sum_{i=1}^{n_1} (y_{1i} - \bar{y}_1)^2 + \sum_{i=1}^{n_2} (x_{2i} - \bar{x}_2)^2 + \sum_{i=1}^{n_2} (y_{2i} - \bar{y}_2)^2,$$

where x_{1i}, y_{1i} $(i = 1, 2, \cdots, n_1)$ are the pairs of observations in the first sample and x_{2i}, y_{2i} $(i = 1, 2, \cdots, n_2)$ are the pairs of observations in the second sample. $S'_{r'}/\sigma^2$ is distributed according to the χ^2-distribution with $(2n_1 + 2n_2 - 4)$ degrees of freedom because it is the sum of quantities independently distributed as χ^2. Then

$$R' = \frac{n_1 n_2 r'^2}{2(n_1 + n_2)\sigma^2} \Big/ \frac{S'_{r'}}{(2n_1 + 2n_2 - 4)\sigma^2} = \frac{n_1 n_2 (n_1 + n_2 - 2) r'^2}{(n_1 + n_2) S'_{r'}}$$

has the F-distribution with 2 and $(2n_1 + 2n_2 - 4)$ degrees of freedom. This fact yields us a significance test for the hypothesis that both the means of the x variates and the means of the y variates for the two populations are the same. We can also set up confidence regions for $\mu_1 - \mu_2$ and $\nu_1 - \nu_2$.

Now let us consider a sample from a normal bivariate population with means μ and ν, variances σ_x^2 and σ_y^2 and correlation coefficient ρ. Suppose $\gamma = \sigma_x^2/\sigma_y^2$ and ρ are known. The transformation

$$x = \frac{\sqrt{1+\rho}\,x' + \sqrt{1-\rho}\,y'}{\sqrt{2}}, \tag{8}$$

$$y = \frac{\sqrt{1+\rho}\,x' - \sqrt{1-\rho}\,y'}{\sqrt{2}\,\gamma},$$

gives us the variates x' and y' which are distributed independently and with variance σ_x^2. Applying the results above we see that

$$\begin{aligned} R &= n(n-1)\,\frac{(\bar{x}' - \mu')^2 + (\bar{y}' - \nu')^2}{\sum_{i=1}^{n} (x'_i - \bar{x}')^2 + \sum_{i=1}^{n} (y'_i - \bar{y}')^2} \\ &= n(n-1)\,\frac{(\bar{x} - \mu)^2 - 2\rho\sqrt{\gamma}\,(\bar{x} - \mu)(\bar{y} - \nu) + \gamma(\bar{y} - \nu)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2 - 2\rho\sqrt{\gamma}\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) + \gamma\sum_{i=1}^{n} (y_i - \bar{y})^2} \end{aligned} \tag{9}$$

has the F-distribution with 2 and $2n - 2$ degrees of freedom. From this we derive significance tests, fiducial radii, and confidence regions as before.

The above distributions, significance tests, and confidence regions are easily generalized to multivariate normal distributions. Suppose we have a sample of n k-tuples of observations $\{x_{i\alpha}\}$ $(i = 1, 2, \cdots, k; \alpha = 1, 2, \cdots, n)$ from a k-variate normal distribution. Let the expected value of each variate x_i be zero $(i = 1, 2, \cdots, k)$, the variance of each variate be σ^2 and each correlation coefficient be zero. Then

$$R'' = \frac{n(n-1)\sum_{i=1}^{k} \bar{x}_i^2}{\sum_{i=1}^{k}\sum_{\alpha=1}^{n} (x_{i\alpha} - \bar{x}_i)^2} \tag{10}$$

has the F-distribution with k and $k(n - 1)$ degrees of freedom. Significance tests, confidence regions, and fiducial radii follow from this fact.

3. Linear Regression. If one has a sample of n pairs of observations $(x_1, y_1; x_2, y_2; \cdots; x_n, y_n)$ from a normal bivariate population and wishes to fit a straight line to the scatter of sample points, one fits the line in such a way that the sum of squares of distances from the sample points to the line is a minimum ("error in both variates").

It is easily shown that this line goes through the point whose coordinates are the sample means $(\bar{x}, \bar{y}.)$ If the slope of a line through $(\bar{x}, \bar{y})$ is $\tan \theta$, the distance from a sample point (x_i, y_i) to the line is $(x_i - \bar{x}) \sin \theta - (y_i - \bar{y}) \cos \theta$. The sum of squares of distances from sample points to the line is

$$\sin^2 \theta \, S_x - 2 \sin \theta \cos \theta \, S_{xy} + \cos^2 \theta \, S_y ,$$

where

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}).$$

If we minimize the above expression with respect to θ we find

$$b = \tan \theta = \frac{S_y - S_x \pm \sqrt{(S_y - S_x)^2 + 4S_{xy}^2}}{2S_{xy}} . \tag{11}$$

Using the plus sign gives us S_p, the minimum sum of squared distances; using the minus sign gives us S_a, the maximum sum of squared distances. (The latter value of $\tan \theta$ is the negative reciprocal of the former.)

S_p is the sum of squared distances perpendicular to the regression line and S_a is the sum of squared distances along the regression line. The sum $S_p + S_a$ is equal to $S_x + S_y$ which is the sum of squares of distances from the sample points to the point $\bar{x}, \bar{y}$. We have thus decomposed $S_x + S_y$ into two components, one perpendicular to the regression line and the other along the regression line.

The joint distribution of S_p and S_a may be derived from the Wishart distribution of the sums of squares and cross products,[1]

$$(12) \qquad \frac{1}{4\pi\sigma^{2n-2}\Gamma(n-2)} \begin{vmatrix} S_x & S_{xy} \\ S_{xy} & S_y \end{vmatrix}^{\frac{1}{2}(n-4)} e^{-\frac{1}{2}(S_x+S_y)/\sigma^2}$$

Let us make the transformation

$$S_x = \cos^2\theta\, S_a + \sin^2\theta\, S_p,$$

$$S_y = \sin^2\theta\, S_a + \cos^2\theta\, S_p,$$

$$S_{xy} = \sin\theta\cos\theta\,(S_a - S_p).$$

The value of θ corresponds to the plus sign in (11). We find

$$S_x + S_y = S_p + S_a,$$

$$\begin{vmatrix} S_x & S_{xy} \\ S_{xy} & S_y \end{vmatrix} = S_p S_a.$$

The Jacobian of the transformation is $(S_a - S_p)$. Using these relations in (12) and integrating out θ we derive the distribution of S_a and S_p

$$(13) \qquad \frac{1}{4\sigma^6\Gamma(n-2)} \left(\frac{S_a S_p}{\sigma^4}\right)^{\frac{1}{2}(n-4)} e^{-\frac{1}{2}(S_a+S_p)/\sigma^2}\,(S_a - S_p).$$

It can be shown that S_a and S_p are the characteristic roots of the sample variance-covariance matrix. The distribution (13) of the characteristic roots of a variance-covariance matrix when the population correlation coefficient is zero and the variances are equal has been demonstrated by P. L. Hsu [4].

As a test of correlation (i.e., test of significance of the regression coefficient) we propose using the ratio

$$F' = S_a/S_p.$$

This ratio is the maximum ratio of the sum of squared deviations in one direction to the sum of squared deviations in the perpendicular direction. It is intuitively evident that this ratio is probably near unity if the null hypothesis is true, that is, if the variances are equal and the correlation is zero. If the correlation is not zero then the ratio is likely to be large.

From (13) we can deduce the distribution of F' by transforming variables and integrating out the extraneous one. This procedure yields us as the distribution of F'

$$(n-2)2^{n-3}F'^{\frac{1}{2}(n-4)}(F'+1)^{-(n-1)}(F'-1).$$

If we make the transformation

$$F' = e^{2z'},$$

[1] This distribution is equivalent to Fisher's distribution of the sample variances and correlation coefficient when the population correlation coefficient is zero.

we find the probability element of z' to be

$$(n - 2)(\cosh z')^{-(n-1)}\, d(\cosh z')$$

After integrating we see the cumulative distribution of z' is

$$1 - (\cosh z')^{-(n-2)}.$$

Critical values of z' for various levels of significance may be determined from a table of hyperbolic cosines. Table I gives some values of z' and the corresponding values of F'.

TABLE I

Percentage points for the z' (or F') distribution

n	z'					F'				
	$P_{.20}$	$P_{.10}$	$P_{.05}$	$P_{.01}$	$P_{.001}$	$P_{.20}$	$P_{.10}$	$P_{.05}$	$P_{.01}$	$P_{.001}$
3	2.292	2.993	3.688	5.298	7.601	98.0	398	1600	40,000	4,000,000
4	1.444	1.818	2.178	2.993	4.144	17.9	38.0	78.0	398	4,000
5	1.130	1.402	1.656	2.216	2.993	9.59	16.5	27.4	84.2	398
6	.958	1.178	1.381	1.818	2.412	6.79	10.6	15.8	38.0	124
7	.846	1.035	1.207	1.572	2.059	5.43	7.92	11.2	23.2	61.4
8	.766	.933	1.084	1.402	1.818	4.63	6.47	8.74	16.5	38.0
9	.704	.856	.992	1.276	1.643	4.09	5.55	7.28	12.7	26.8
10	.656	.796	.920	1.178	1.509	3.71	4.91	6.30	10.6	20.5
11	.616	.746	.862	1.100	1.402	3.43	4.45	5.61	9.02	16.5
12	.583	.705	.813	1.035	1.314	3.21	4.10	5.09	7.92	13.9
13	.554	.670	.772	.980	1.241	3.03	3.82	4.68	7.10	12.0
14	.530	.639	.736	.933	1.178	2.89	3.59	4.36	6.47	10.6
15	.508	.613	.705	.892	1.124	2.76	3.41	4.10	6.00	9.47
20	.429	.517	.593	.746	.993	2.36	2.81	3.27	4.45	6.47
25	.378	.455	.522	.654	.814	2.13	2.48	2.84	3.70	5.10
30	.342	.411	.471	.589	.732	1.98	2.28	2.57	3.25	4.32
40	.293	.352	.402	.502	.621	1.80	2.02	2.23	2.73	3.47
60	.237	.284	.324	.404	.498	1.61	1.76	1.91	2.24	2.71
120	.165	.198	.226	.281	.345	1.39	1.49	1.57	1.75	2.00

The use of F' has been suggested here to test the hypothesis that the population correlation coefficient is zero when it is known that the variances of the two variates are the same, or, more generally, when the ratio of the two variances is known. This gives a test of significance of the regression coefficient when there is error in both variates if the ratio of the variances is known. The test arises from intuitive considerations. F' can also be used to test the hypothesis that $\rho = 0$ *and* $\sigma_x^2 = \sigma_y^2$ (H_4 in Hsu's paper). C. T. Hsu [5] and J. W. Mauchly [6] have shown that the likelihood ratio criterion for this hypothesis is

$$\lambda = \left[\frac{2(S_x S_y - S_{xy}^2)}{(S_x + S_y)^2}\right]^{\frac{1}{2}n} = \left[\frac{2F'}{(F' + 1)^2}\right]^{\frac{1}{2}n}.$$

If we set the normal distribution function equal to a constant, we determine a contour ellipse in the x, y — plane. Since these ellipses of constant probability density are circles when $\rho = 0$ and $\sigma_x^2 = \sigma_y^2$, Mauchly calls the test a test of circularity. The same procedure as used to test whether these ellipses are circles can be used to test whether the ellipses have major axes in a certain direction and with a specified ratio of lengths of axes. Suppose we wish to test the hypothesis that the major axis is inclined to the x axis at an angle θ and that the ratio of lengths of the major axis to the minor axis is k. This is equivalent to the hypothesis that $\rho = \rho_0$ and $\sigma_x^2 = \gamma_0\sigma_y^2$. To do this we rotate coordinate axes of the variables of the distribution (hence changing coordinates of all sample points) through θ and change the scale of one of the new variables by the factor of k. The transformation is

$$x = kx' \cos\theta - y' \sin\theta,$$

$$y = kx' \sin\theta + y' \cos\theta.$$

In terms of x', y' the null hypothesis is $\rho' = 0$, $\sigma_{x'}^2 = \sigma_{y'}^2$, and one proceeds as above. Of course, if γ_0 is known then this method can be used to test the null hypothesis that $\rho = \rho_0$.

4. Illustrative Example. An application of the formulae given above may be illustrated from the data in Table II, which gives two sets of electrical conductivity measurements at different field strengths. The assumption that the two variances are equal is thus reasonable.

Table of Pairs of Observations of Electrical Conductivity

x_i	y_i	x_i	y_i
5.0	5.1	5.5	5.1
7.4	7.0	5.3	5.0
7.0	7.7	4.7	4.4
8.8	7.7	8.6	7.1
7.8	6.8	7.5	7.3
5.1	5.5	5.6	6.3
6.6	7.4	7.4	6.5
8.8	7.7		

Is it reasonable to regard x and y as being independently distributed in the population on the basis of these data?

The sums of squares and cross products of deviations from the means and the calculated slope are:

$$S_x = 29.40, \qquad S_{xy} = 19.99,$$

$$S_y = 18.04, \qquad b = 0.7554.$$

The maximized variance ratio is:

$$F' = \frac{S_x + 2bS_{xy} + b^2 S_y}{b^2 S_x - 2bS_{xy} + S_y} = \frac{69.89}{4.615} = 15.15.$$

$$z' = \tfrac{1}{2}\ln F' = 1.36.$$

Comparing with Table I for $n = 15$ we find this value of z' very highly significant (probability less than 0.001), and at this probability level and on basis of our data, x and y cannot be considered to be independent in the population.

Since the regression is significant, it becomes of interest to compute the calculated points X_i and Y_i which fall on the regression line

$$Y = 1.35 + 0.7554\,X,$$

corresponding to each observed point x_i, y_i. They are obtained from these equations

$$Y_i = \bar{y} + \frac{b}{1+b^2}(x_i - \bar{x}) + \frac{b^2}{1+b^2}(y_i - \bar{y})$$

$$= .481x_i + .363y_i + .86,$$

$$X_i = \bar{x} + \frac{1}{1+b^2}(x_i - \bar{x}) + \frac{b}{1+b^2}(y_i - \bar{y})$$

$$= .637x_i + .481y_i - .65.$$

The minimized sum of squared deviations from the regression line (i.e., squared distances between observed and calculated points) is the denominator of the expression for F' divided by the factor $(1 + b^2)$,

$$4.615/.5706 = 2.64.$$

It should perhaps be pointed out that the tests of the means described in the first part of this paper are no longer applicable since we do not know the population correlation coefficient.

REFERENCES

[1] G. W. Snedecor, *Statistical Methods*, Iowa State College Press (1940), pp. 184–187.

[2] S. Kolodziejczyk, "On an important class of statistical hypotheses," *Biometrika*, Vol. 27 (1935), pp. 161–190.

[3] R. A. Fisher, "The fiducial argument in statistical inference," *Annals of Eugenics*, Vol. 6 (1935), pp. 391–398.

[4] P. L. Hsu, "On the distribution of roots of certain determinantal equations," *Annals of Eugenics*, Vol. 9 (1939), pp. 250–258.

[5] C. T. Hsu, "On samples from a normal bivariate population," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 410–426.

[6] J. W. Mauchly, "Significance test for sphericity of a normal n-variate distribution," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 204–209.

SYMMETRIC TESTS OF THE HYPOTHESIS THAT THE MEAN OF ONE NORMAL POPULATION EXCEEDS THAT OF ANOTHER

By Herbert A. Simon

Illinois Institute of Technology

1. Introduction. One of the most commonly recurring statistical problems is to determine, on the basis of statistical evidence, which of two samples, drawn from different universes, came from the universe with the larger mean value of a particular variate. Let M_y be the mean value which would be obtained with universe (Y) and M_x be the mean value which would be obtained with universe (X). Then a test may be constructed for the hypothesis[1] $M_y \geq M_x$.

If $x_1, \cdots, x_n$ are the observed values of the variate obtained from universe (X), and $y_1, \cdots, y_n$ are the observed values obtained from universe (Y), then the sample space of the points $E: (x_1, \cdots, x_n; y_1, \cdots, y_n)$ may be divided into three regions ω_0, ω_1, and ω_2. If the sample point falls in the region ω_0, the hypothesis $M_y \geq M_x$ is accepted; if the sample point falls in the region ω_1, the hypothesis $M_y \geq M_x$ is rejected; if the sample point falls in the region ω_2, judgment is withheld on the hypothesis. Regions ω_0, ω_1, and ω_2 are mutually exclusive and, together, fill the entire sample space. Any such set of regions ω_0, ω_1, and ω_2 defines a test for the hypothesis $M_y \geq M_x$.

In those cases, then, where the experimental results fall in the region ω_2, the test leads to the conclusion that there is need for additional data to establish a result beyond reasonable doubt. Under these conditions, the test does not afford any guide to an unavoidable or non-postponable choice. In the application of statistical findings to practical problems it often happens, however, that judgment can not be held in abeyance—that *some* choice must be made, even at a risk of error. For example, when planting time comes, a choice must be made between varieties (X) and (Y) of grain even if neither has been conclusively demonstrated, up to that time, to yield a larger crop than the other. It is the purpose of this paper to propose a criterion which will always permit a choice between two experimental results, that is, a test in which the regions ω_0 and ω_1 fill the entire sample space. In the absence of a region ω_2, any observed result is interpreted as a definite acceptance or rejection of the hypothesis tested.

2. General characteristics of the criterion. Let us designate the hypothesis $M_y \geq M_x$ as H_0 and the hypothesis $M_x > M_y$ as H_1. Then a pair of tests, T_0 and T_1, for H_0 and H_1 respectively must, to suit our needs, have the following properties:

(1) The regions ω_{00} (ω_{00} is the region of acceptance for H_0, ω_{10} the region of rejection for H_0; ω_{01} and ω_{11} the corresponding regions for H_1) and ω_{11} must

[1] This paper presupposes a familiarity with the theory of testing statistical hypotheses as set forth by J. Neyman and E. S. Pearson [1].

coincide; as must the regions ω_{10} and ω_{01}. This correspondence means that when H_0 is accepted, H_1 is rejected, and vice versa. Hence, the tests T_0 and T_1 are identical, and we shall hereafter refer only to the former.

(2) There must be no regions ω_{20} and ω_{21}. This means that judgment is never held in abeyance, no matter what sample is observed.

(3) The regions ω_{00} and ω_{10} must be so bounded that the probability of accepting H_1 when H_0 is true (error of the first kind for T_0) and the probability of accepting H_0 when H_1 is true (error of the second kind for T_0) are, in a certain sense, minimized. Since H_0 and H_1 are composite hypotheses, the probability that a test will accept H_1 when H_0 is true depends upon which of the simple hypotheses that make up H_0 is true.

Neyman and Pearson [2] have proposed that a test, T_α for a hypothesis be termed *uniformly more powerful* than another test, T_β, if the probability for T_α of accepting the hypothesis if it is false, or the probability of rejecting it if it is true, does not exceed the corresponding probability for T_β no matter which of the simple hypotheses is actually true. Since there is no test which is uniformly more powerful than all other possible tests, it is usually required that a test be uniformly most powerful (UMP) among the members of some specified class of tests.

3. A symmetric test when the two universes have equal standard deviations. Let us consider, first, the hypothesis $M_y \geq M_x$ where the universes from which observations of varieties (X) and (Y), respectively, are drawn are normally distributed universes with equal standard deviations, σ, and means M_x and M_y respectively. Let us suppose a sample drawn of n random observations from the universe of variety (X) and a sample of n independent and random observations from the universe of (Y). The probability distribution of points in the sample space is given by

$$(1) \qquad p(x_1, \cdots, x_n; y_1, \cdots, y_n) = (2\pi\sigma^2)^{-n} e^{-\frac{1}{2\sigma^2}\left[\sum_i (x_i - M_x)^2 + \sum_i (y_i - M_y)^2\right]}.$$

In testing the hypothesis $M_y \geq M_x$, there is a certain symmetry between the alternatives (X) and (Y). If there is no *a priori* reason for choosing (X) rather than (Y), and if the sample point E_1: $(a_1, \cdots, a_n; b_1, \cdots, b_n)$ falls in the region of acceptance of H_0: then the point E_2: $(b_1, \cdots, b_n; a_1, \cdots, a_n)$ should fall in the region of acceptance of H_1. That is, if E_1 is taken as evidence that $M_y \geq M_x$; then E_2 can with equal plausibility be taken as evidence that $M_x \geq M_y$.

Any test such that E_1: $(a_1, \cdots, a_n; b_1, \cdots, b_n)$ lies in ω_0 whenever E_2: $(b_1, \cdots, b_n; a_1, \cdots, a_n)$ lies in ω_1 and vice versa, will be designated a symmetric test of the hypothesis $M_y \geq M_x$. Let Ω be the class of symmetric tests of H_0. If T_α is a member of Ω, and is uniformly more powerful than every other T_β which is a member of Ω, then T_α *is the uniformly most powerful symmetric test of* H_0.

The hypothesis $M_y \geq M_x$ possesses a UMP symmetric test. This may be shown as follows. From (1), the ratio can be calculated between the proba-

bility densities at the sample points $E:(x_1, \cdots, x_n; y_1, \cdots, y_n)$ and E': $(y_1, \cdots, y_n; x_1, \cdots, x_n)$. We get

$$\frac{p(E)}{p(E')} = \exp\left\{\frac{n}{\sigma^2}(\bar{x} - \bar{y})(M_x - M_y)\right\}, \tag{2}$$

where

$$\bar{x} = \frac{1}{n}\sum_i x_i, \qquad \bar{y} = \frac{1}{n}\sum_i y_i.$$

Now the condition $p(E) > p(E')$ is equivalent to $\frac{n}{\sigma^2}(\bar{x} - \bar{y})(M_x - M_y) > 0$. Hence $p(E) > p(E')$ whenever $(\bar{x} - \bar{y})$ has the same sign as $(M_x - M_y)$.

Now for any symmetric test, if E lies in ω_0, E' lies in ω_1, and vice versa. Suppose that, in fact, $M_y > M_x$. Consider a symmetric test, T_α whose region ω_0 contains a sub-region ω_{0U} (of measure greater than zero) such that $\bar{y} < \bar{x}$ for every point in that sub-region. Then for every point E' in ω_{0U}, $p(E') < p(E)$. Hence, a more powerful test, T_β could be constructed which would be identical with T_α, except that ω_{1U}, the sub-region symmetric to ω_{0U}, would be interchanged with ω_{0U} as a portion of the region of acceptance for H_0. Therefore, a test such that ω_0 contained all points for which $\bar{y} > \bar{x}$, and no others, would be a UMP symmetric test. This result is independent of the magnitude of $(M_x - M_y)$ provided only $M_y \geq M_x$. We conclude that $\bar{y} > \bar{x}$ *is a uniformly most powerful symmetric test for the hypothesis* $M_y > M_x$.

The probability of committing an error with the UMP symmetric test is a simple function of the difference $|M_y - M_x|$. The exact value can be found by integrating (1) over the whole region of the sample space for which $\bar{y} < \bar{x}$. There is no need to distinguish errors of the first and second kind, since an error of the first kind with T_0 is an error of the second kind with T_1, and vice versa. The probability of an error is one half when $M_x = M_y$, and in all other cases is less than one half.

4. Relation of UMP symmetric test and test which is UMP of tests absolutely equivalent to it. Neyman and Pearson [2] have shown the test $\bar{y} - \bar{x} > k$ to be UMP among the tests absolutely equivalent to it, for the hypothesis $M_y \geq M_x$. They have defined a class of tests as absolutely equivalent if, for each simple hypothesis in H_0, the probability of an error of the first kind is exactly the same for all the tests which are members of the class. If k be set equal to zero, $\bar{y} > \bar{x}$, and their test reduces to the UMP symmetric test. What is the relation between these two classes of tests?

If T_α be the UMP symmetric test, then it is clear from Section 2 that there is no other symmetric test, T_β, which is absolutely equivalent to T_α. Hence Ω, the class of symmetric tests, and Λ, the class of tests aboslutely equivalent to T_α, have only one member in common—the test T_α itself. Neyman and Pearson have shown T_α to be the UMP test of Λ, while the results of Section 4 show T_α to be the UMP test of Ω.

5. Justification for employing a symmetric test. In introducing Section 3, a heuristic argument was advanced for the use of a symmetric, rather than an asymmetric test for the hypothesis $M_y \geq M_x$. This argument will now be given a precise interpretation in terms of probabilities.

Assume, not a single experiment for testing the hypothesis $M_y \geq M_x$, but a series of similar experiments. Suppose a judgment to be formed independently on the basis of each experiment as to the correctness of the hypothesis. Is there any test which, if applied to the evidence in each case, will maximize the probability of a correct judgment in that experiment? Such a test can be shown to exist, providing one further assumption is made: that if any criterion be applied *prior* to the experiment to test the hypothesis $M_y \geq M_x$, the probability of a correct decision will be one half. That is, it must be assumed that there is no evidence which, prior to the experiment, will permit the variety with the greater yield to be selected with greater-than-chance frequency.

Consider now any asymmetric test for the hypothesis H_0—that is, any test which is not symmetric. The criterion $\bar{y} - \bar{x} > k$, where $k > 0$, is an example of such a test. Unlike a symmetric test, an asymmetric test may give a different result if applied as a test of the hypothesis H_0 than if applied as a test of the hypothesis H_1. For instance, a sample point such that $\bar{y} - \bar{x} = \epsilon$, where $k > \epsilon > 0$, would be considered a rejection of H_0 and acceptance of H_1 if the above test were applied to H_0; but would be considered a rejection of H_1 and an acceptance of H_0 if the test were applied to H_1. Hence, before an asymmetric test can be applied to a problem of dichotomous choice—a problem where H_0 *or* H_1 must be determinately selected—a decision must be reached as to whether the test is to be applied to H_0 or to H_1. This decision cannot be based upon the evidence of the sample to be tested—for in this case, the complete test, which would of course include this preliminary decision, would be symmetric by definition.

Let H_c be the correct hypothesis (H_0 or H_1, as the case may be) and let H_* be the hypothesis to which the asymmetric test is applied. Since by assumption there is no prior evidence for deciding whether H_c is H_0 or H_1, we may employ any random process for deciding whether H_* is to be identified with H_0 or H_1. If such a random selection is made, it follows that the probability that H_c and H_* are identical is one half.

We designate as the region of asymmetry of a test the region of points E_1: $(a_1, \cdots, a_n; b_1, \cdots, b_n)$ and E_2: $(b_1, \cdots, b_n; a_1, \cdots, a_n)$ of aggregate measure greater than zero such that E_1 and E_2 both fall in ω_0 or both fall in ω_1. Suppose ω_{0a} and ω_{0b} are a particular symmetrically disposed pair of subregions of the region of asymmetry, which fall in ω_0 of a test T_0. Suppose that, for every point, E_1, in ω_{0a}, $\bar{b} > \bar{a}$, and that ω_{0a} and ω_{0b} are of measure greater than zero. The sum of the probabilities that the sample point will fall in ω_{0a} or ω_{0b} is exactly the same whether H_c and H_* are the same hypothesis or are contradictory hypotheses. In the first case H_c will be accepted, in the second case H_c will be rejected. These two cases are of equal probability, hence there is a probability

of one half of accepting or rejecting H_c if the sample point falls in the region of asymmetry of T_0. But from equation (2) of Section 2 above, we see that if the subregions ω_{0a} and ω_{0b} had been in a region of symmetry, and if ω_{0a} had been in ω_0, the probability of accepting H_c would have been greater than the probability of rejecting H_c.

Hence, if it is determined by random selection to which of a pair of hypotheses an asymmetric test is going to be applied, the probability of a correct judgment with the asymmetric test will be less than if there were substituted for it the UMP symmetric test. It may be concluded that the UMP symmetric test is to be preferred unless there is prior evidence which permits a tentative selection of the correct hypothesis with greater-than-chance frequency.

6. Symmetric test when standard deviations of universes are unequal. Thus far, we have restricted ourselves to the case where $\sigma_x = \sigma_y$. Let us now relax this condition and see whether a UMP symmetric test for $M_y \geq M_x$ exists in this more general case.

We now have for the ratio of $p(E)$ to $p(E')$:

$$(3) \quad \frac{p(E)}{p(E')} = \exp\left\{-\frac{n}{2\sigma_x^2\sigma_y^2}[(\sigma_y^2 - \sigma_x^2)(\mu_x - \mu_y) - 2(\sigma_y^2 M_x - \sigma_x^2 M_y)(\bar{x} - \bar{y})]\right\},$$

where

$$\mu_x = \sum_i x_i^2/n, \qquad \mu_y = \sum_i y_i^2/n.$$

Even if σ_y and σ_x are known, which is not usually the case, there is no UMP symmetric test for the hypothesis $M_y \geq M_x$. From (3), the symmetric critical region which has the lowest probability of errors of the first kind for the hypothesis $(M_y = k_1 ; M_x = k_2 ; k_1 > k_2)$ is the set of points E such that:

$$(4) \qquad (\sigma_y^2 - \sigma_x^2)(\mu_x - \mu_y) - 2(\sigma_y^2 k_2 - \sigma_x^2 k_1)(\bar{x} - \bar{y}) > 0.$$

Since this region is not the same for all values of k_1 and k_2 such that $k_1 > k_2$, there is no UMP symmetric region for the composite hypothesis $M_y \geq M_x$. This result holds, *a fortiori* when σ_y and σ_x are not known.

If there is no UMP symmetric test for $M_y \geq M_x$ when $\sigma_y \neq \sigma_x$, we must be satisfied with a test which is UMP among some class of tests more restricted than the class of symmetric tests. Let us continue to restrict outselves to the case where there are an equal number of observations, in our sample, of (X) and of (Y). Let us pair the observations x_i, y_i, and consider the differences $u_i = x_i - y_i$. Is there a UMP test among the tests which are symmetric with respect to the u_i's for the hypothesis that $M_y - M_x = -U \geq 0$? By a symmetric test in this case we mean a test such that whenever the point $(u_1, \cdots, u_n)$ falls into region ω_0, the point $(-u_1, \cdots, -u_n)$ falls into region ω_1.

If x_i and y_i are distributed normally about M_x and M_y with standard deviations σ_x and σ_y respectively, then u_i will be normally distributed about $U =$

$M_x - M_y$ with standard deviation $\sigma_u = \sqrt{\sigma_x^2 + \sigma_y^2}$. The ratio of probabilities for the sample points $E_v: (u_1, \cdots, u_n)$ and $E_v': (-u_1, \cdots, -u_n)$ is given by:

$$\frac{p(E_v)}{p(E_v')} = \exp\left\{\frac{-2n}{\sigma_u^2}\,\bar{u}U\right\}, \tag{5}$$

where

$$\bar{u} = \frac{1}{n}\sum_i u_i .$$

Hence, $p(E_v) > p(E_v')$ whenever $\bar{u}$ has the same sign as U. Therefore, by the same process of reasoning as in Section 2, above, we may show that $\bar{u} \leq 0$ is a UMP test among tests symmetric in the sample space of the u's for the hypothesis $U \leq 0$.

It should be emphasized that Ω_{su}, the class of symmetric regions in the space of $E_v: (u_1 \cdots u_n)$, is far more restricted than Ω_s, the class of symmetric regions in the sample space of $E: (x_1 \cdots x_n ; y_1 \cdots y_n)$. In the latter class are included all regions such that:

(A) $E: (a_1, \cdots, a_n ; b_1, \cdots, b_n)$ falls in ω_0 whenever $E: (b_1, \cdots, b_n ; a_1, \cdots, a_n)$ falls in ω_1. Members of class Ω_{su} satisfy this condition together with the further condition:

(B) For all possible sets of n constants $k_1, \cdots, k_n$, $E: (x_1 + k_1, \cdots, x_n + k_n ; y_1 + k_1, \cdots, y_n + k_n)$ falls in ω_0 whenever $E: (x_1, \cdots, x_n ; y_1, \cdots, y_n)$ falls in ω_0. When $\sigma_y \neq \sigma_x$, a UMP test for $M_y \geq M_x$ with respect to the symmetric class Ω_{su} exists, but a UMP test with respect to the symmetric class Ω_s does not exist.

REFERENCES

[1] J. NEYMAN and E. S. PEARSON, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. Roy. Soc.*, Series A, 702, Vol. 231 (1933), pp. 289–337.

[2] J. NEYMAN and E. S. PEARSON, "The testing of statistical hypotheses in relation to probabilities *A Priori*," *Proc. Camb. Phil. Soc.*, Vol. 29 (1933), pp. 492–510.

ON INDICES OF DISPERSION

By Paul G. Hoel

University of California, Los Angeles

1. Introduction. In biological sciences the index of dispersion for the binomial and Poisson distributions is very useful for testing homogeneity of certain types of data. For example, the dilution technique in making blood counts finds it useful. Recently there have been attempts to use it to determine allergies by observing the change in the blood count after allergic foods have been taken. Here the sample may consist of only a few readings; consequently it is important to know how accurate this index is when applied to small samples. After inspecting the application of the Poisson index to such counts, I was surprised to see the lack of agreement with theory. At first it appeared that the fault lay with the chi-square approximation which is used on this index, but later it was clear that the assumption of a basic Poisson distribution was at fault. It now appears that statisticians will need to be careful about citing blood counts as examples of data following a Poisson distribution.

This paper is the result of investigating the accuracy of the chi-square approximation for the distribution of these indices. Previous work on this problem seems to have consisted in some sampling experiments [1] for small values of the parameters involved, and in some theoretical work [2] in which the sampling distribution is considered only for a fixed sample mean. Although sampling distributions ordinarily differ very little from the distributions obtained by assuming the mean of the sample fixed, for small degrees of freedom the difference may be appreciable and therefore requires investigation. In this paper the accuracy of the chi-square approximation is investigated by finding expressions for the descriptive moments of the distribution which are correct to terms of order N^{-3}. These expressions are obtained by means of Fisher's semi-invariant technique.

2. Moments of the distribution. Employing Fisher's notation [3], let the binomial index of dispersion be denoted by z, then z may be written as:

$$z = \frac{\Sigma(x - \bar{x})^2}{\bar{x}\left(1 - \frac{\bar{x}}{n}\right)} = \frac{(N-1)k_2}{k_1\left(1 - \frac{k_1}{n}\right)} = \frac{N-1}{\kappa_1\left(1 - \frac{\kappa_1}{n}\right)} \frac{k_2}{\left(1 + \frac{k_1 - \kappa_1}{\kappa_1}\right)\left(1 - \frac{k_1 - \kappa_1}{n - \kappa_1}\right)}.$$

Letting $w = k_1 - \kappa_1$, $y = k_2$, $a = n \dot{-} \kappa_1$, $b = \dfrac{N-1}{\kappa_1\left(1 - \frac{\kappa_1}{n}\right)}$, z may be ex-

panded as follows:

$$
\begin{aligned}
z &= \frac{by}{\left(1+\frac{w}{\kappa_1}\right)\left(1-\frac{w}{a}\right)} \\
&= by\left\{1-\frac{w}{\kappa_1}+\frac{w^2}{\kappa_1^2}-\cdots\right\}\left\{1+\frac{w}{a}+\frac{w^2}{a^2}+\cdots\right\} \\
&= by\left\{1+w\left(\frac{1}{a}-\frac{1}{\kappa_1}\right)+w^2\left(\frac{1}{a^2}-\frac{1}{a\kappa_1}+\frac{1}{\kappa_1^2}\right)+\cdots\right\} \\
&= b\{y+c_1wy+c_2w^2y+c_3w^3y+\cdots\},
\end{aligned}
$$

where the definition of c_i is obvious. As will be seen later, these expansions are valid for obtaining the expected values of powers of z; hence

$$
\begin{aligned}
E(z) &= b\ \{\mu_{01}+c_1\mu_{11}+c_2\mu_{21}+\cdots\} \\
E(z^2) &= b^2\{\mu_{02}+2c_1\mu_{12}+(2c_2+c_1^2)\mu_{22}+(2c_3+2c_2c_1)\mu_{32}+\cdots\} \\
E(z^3) &= b^3\{\mu_{03}+3c_1\mu_{13}+(3c_2+3c_1^2)\mu_{23}+(3c_3+6c_2c_1+c_1^3)\mu_{33}+\cdots\} \\
E(z^4) &= b^4\{\mu_{04}+4c_1\mu_{14}+(4c_2+6c_1^2)\mu_{24}+(4c_3+12c_2c_1+4c_1^3)\mu_{34}+\cdots\}.
\end{aligned} \tag{1}
$$

Since only the first four moments of z are to be found, it will be necessary to evaluate the μ_{ij} for $j = 1, 2, 3, 4$ and for $i = 0, 1, 2, \cdots$ as far as necessary to give the desired degree of accuracy.

First consider the relation between the moments μ_{ij} and the semi-invariants κ_{ij} which are defined in terms of the μ_{ij} by the following formal identity in t and τ.

$$
e^{\frac{\kappa_{10}t+\kappa_{01}\tau}{1!}+\frac{\kappa_{20}t^2+2\kappa_{11}t\tau+\kappa_{02}\tau^2}{2!}+\cdots}=1+\frac{\mu_{10}t+\mu_{01}\tau}{1!}+\frac{\mu_{20}t^2+2\mu_{11}t\tau+\mu_{02}\tau^2}{2!}+\cdots.
$$

Differentiating both sides with respect to t and replacing the exponential factor by the right member gives an identity which is convenient for evaluating the μ_{i0}. Differentiating both sides with respect to τ and making the same replacement gives an identity which is convenient for evaluating the μ_{ij} for $j > 0$.

These identities express μ_{ij} as a sum of products of κ's and μ's, each such product being of total degree i and j in its subscripts. By repeated substitution, μ_{ij} can be expressed as a sum of products of κ's only. From Fisher's formulas

each such semi-invariant, κ_{rs}, can be expressed as a sum of products of semi-invariants of the basic distribution, each term of which sum is of order $N^{-(r+s-1)}$ in N. Hence it follows that the lowest order term, or at least one of the lowest order terms, in N in the expression for μ_{ij} will be a term with the maximum number of κ factors. Since the κ_{rs} of lowest degree in subscripts are κ_{10} and κ_{01}, the term with the maximum number of κ factors will be the term in $\kappa_{10}^{i}\kappa_{01}^{j}$. However, since $w = k_1 - \kappa_1$ has a zero mean value, $\mu_{10} = \kappa_{10} = 0$; consequently the lowest degree term involving the subscript $i > 0$ is κ_{20} or κ_{11}. As a result, the maximum number of κ factors will be found in the term containing $\kappa_{20}^{\frac{1}{2}i}\kappa_{01}^{j}$ for i even and $\kappa_{20}^{\frac{1}{2}(i-1)}\kappa_{01}^{j-1}\kappa_{11}$ for i odd. These terms are of order $N^{-\frac{1}{2}i}$ and $N^{-\frac{1}{2}(i+1)}$ respectively. Since it is desired to obtain accuracy of order N^{-3}, it therefore will suffice to evaluate μ_{ij} for $i \leq 6$.

The validity of the expansions used in arriving at (1) could now be shown by writing them as partial sums with remainder terms and then showing that the remainder terms are of higher order than N^{-3}.

Neglecting terms of higher order than N^{-3}, the above identities give the following expressions for μ_{ij} for $j = 0, 1, 2$ and $i = 0, 1, \cdots, 6$, with slightly longer expressions for $j = 3$ and 4.

$$\mu_{10} = 0 \qquad \mu_{01} = \kappa_{01}$$
$$\mu_{20} = \kappa_{20} \qquad \mu_{11} = \kappa_{11}$$
$$\mu_{30} = \kappa_{30} \qquad \mu_{21} = \kappa_{21} + \kappa_{01}\mu_{20}$$
$$\mu_{40} = \kappa_{40} + 3\kappa_{20}\mu_{20} \qquad \mu_{31} = \kappa_{31} + 3\kappa_{11}\mu_{20} + \kappa_{01}\mu_{30}$$
$$\mu_{50} = 6\kappa_{30}\mu_{20} + 4\kappa_{20}\mu_{30} \qquad \mu_{41} = 6\kappa_{21}\mu_{20} + 4\kappa_{11}\mu_{30} + \kappa_{01}\mu_{40}$$
$$\mu_{60} = 5\kappa_{20}\mu_{40} \qquad \mu_{51} = 5\kappa_{11}\mu_{40} + \kappa_{01}\mu_{50}$$
$$\mu_{61} = \kappa_{01}\mu_{60}$$

$$\mu_{02} = \kappa_{02} + \kappa_{01}\mu_{01}$$
$$\mu_{12} = \kappa_{12} + \kappa_{11}\mu_{01} + \kappa_{01}\mu_{11}$$
$$\mu_{22} = \kappa_{22} + \kappa_{21}\mu_{01} + \kappa_{02}\mu_{20} + 2\kappa_{11}\mu_{11} + \kappa_{01}\mu_{21}$$
$$\mu_{32} = \kappa_{31}\mu_{01} + 3\kappa_{12}\mu_{20} + 3\kappa_{21}\mu_{11} + \kappa_{02}\mu_{30} + 3\kappa_{11}\mu_{21} + \kappa_{01}\mu_{31}$$
$$\mu_{42} = 6\kappa_{21}\mu_{21} + \kappa_{02}\mu_{40} + 4\kappa_{11}\mu_{31} + \kappa_{01}\mu_{41}$$
$$\mu_{52} = 5\kappa_{11}\mu_{41} + \kappa_{01}\mu_{51}$$
$$\mu_{62} = \kappa_{01}\mu_{61}.$$

The next step is to apply Fisher's formulas expressing the κ_{rs} in terms of the semi-invariants of the basic variable distribution, which in this case is the binomial distribution. In Fisher's notation κ_{rs} would be written as $\kappa(1^r2^s)$, since the variables w and y are respectively k_1, measured from its expected value, and k_2. Applying such formulas, the following expressions for the μ_{i1} and μ_{i2} are obtained, with somewhat longer expressions for the μ_{i3} and μ_{i4}.

$$\mu_{01} = \kappa_2, \qquad \mu_{11} = \frac{\kappa_3}{N}, \qquad \mu_{21} = \frac{\kappa_4}{N^2} + \frac{\kappa_2^2}{N},$$

$$\mu_{31} = \frac{\kappa_5}{N^3} + \frac{4\kappa_3\kappa_2}{N^2}, \qquad \mu_{41} = \frac{7\kappa_4\kappa_2}{N^3} + \frac{4\kappa_3^2}{N^3} + \frac{3\kappa_2^3}{N^2},$$

$$\mu_{51} = \frac{25\kappa_3\kappa_2^2}{N^3}, \qquad \mu_{61} = \frac{15\kappa_2^4}{N^3}$$

$$\mu_{02} = \frac{\kappa_4}{N} + \kappa_2^2\left[\frac{2}{N-1} + 1\right]$$

$$\mu_{12} = \frac{\kappa_5}{N^2} + \frac{2\kappa_3\kappa_2}{N}\left[\frac{2}{N-1} + 1\right] \tag{2}$$

$$\mu_{22} = \frac{\kappa_6}{N^3} + \frac{\kappa_4\kappa_2}{N^2}\left[\frac{4}{N-1} + 3\right] + \frac{2\kappa_3^2}{N^2}\left[\frac{2}{N-1} + 1\right] + \frac{\kappa_2^3}{N}\left[\frac{2}{N-1} + 1\right]$$

$$\mu_{32} = \frac{5\kappa_5\kappa_2}{N^3} + \frac{7\kappa_4\kappa_3}{N^3} + \frac{7\kappa_3\kappa_2^2}{N^2}\left[\frac{2}{N-1} + 1\right]$$

$$\mu_{42} = \frac{16\kappa_4\kappa_2^2}{N^3} + \frac{20\kappa_3^2\kappa_2}{N^3} + \frac{3\kappa_2^4}{N^2}\left[\frac{2}{N-1} + 1\right]$$

$$\mu_{52} = \frac{40\kappa_3\kappa_2^3}{N^3}$$

$$\mu_{62} = \frac{15\kappa_2^5}{N^3}.$$

It is necessary to express these κ's in terms of the parameters of the binomial distribution. Here the κ's are defined by the following formal identity in θ,

$$e^{\kappa_1\theta + \kappa_2\frac{\theta^2}{2!} + \kappa_3\frac{\theta^3}{3!} + \cdots} = (q + pe^{\theta})^n.$$

Taking logarithms, expanding in powers of θ, and equating coefficients of powers of θ, the following expressions are obtained:

$$\kappa_1 = m$$

$$\kappa_2 = mq$$

$$\kappa_3 = mq(q - p)$$

$$\kappa_4 = mq(1 - 6pq)$$

$$\kappa_5 = mq(q - p)(1 - 12pq)$$

$$\kappa_6 = mq(1 - 30pq + 120p^2q^2)$$

$$\kappa_7 = mq(q - p)(1 - 60pq + 360p^2q^2)$$

$$\kappa_8 = mq(1 - 126pq + 1680p^2q^2 - 5040p^3q^3).$$

These values of the κ's are inserted in (2) to give the following expressions for the μ_{i1} and μ_{i2}, with considerably longer expressions for the μ_{i3} and μ_{i4} :

$$\mu_{01} = mq$$

$$\mu_{11} = mq\,(q - p)\,\frac{1}{N}$$

$$\mu_{21} = mq\left(\frac{1 - 6pq}{N^2} + \frac{mq}{N}\right)$$

$$\mu_{31} = mq(q - p)\left(\frac{1 - 12pq}{N^3} + \frac{4mq}{N^2}\right)$$

$$\mu_{41} = m^2q^2\left(\frac{11 - 58pq}{N^3} + \frac{3mq}{N^2}\right)$$

$$\mu_{51} = m^3q^3(q - p)\,\frac{25}{N^3}$$

$$\mu_{61} = m^4q^4\,\frac{15}{N^3}$$

$$\mu_{02} = mq\left(\frac{1 - 6pq}{N} + \frac{2mq}{N - 1} + mq\right)$$

$$\mu_{12} = mq(q - p)\left(\frac{1 - 12pq}{N^2} + \frac{4mq}{N(N - 1)} + \frac{2mq}{N}\right)$$

$$\mu_{22} = mq\left(\frac{1 - 30pq + 120p^2q^2}{N^3} + \frac{8mq(1 - 5pq)}{N^2(N - 1)} + \frac{mq(5 - 26pq)}{N^2} + \frac{2m^2q^2}{N(N - 1)} + \frac{m^2q^2}{N}\right)$$

$$\mu_{32} = m^2q^2(q - p)\left(\frac{12 - 102pq}{N^3} + \frac{14mq}{N^2(N - 1)} + \frac{7mq}{N^2}\right)$$

$$\mu_{42} = m^3q^3\left(\frac{36 - 176pq}{N^3} + \frac{6mq}{N^2(N - 1)} + \frac{3mq}{N^2}\right)$$

$$\mu_{52} = m^4q^4(q - p)\,\frac{40}{N^3}$$

$$\mu_{62} = m^5q^5\,\frac{15}{N^3}.$$

It remains to express the coefficients of (1) in terms of these same parameters. From the definition of c_i, a, and κ_1, it follows that

$$c_i = \frac{\left(\frac{1}{a}\right)^{i+1} + (-1)^i\left(\frac{1}{\kappa_1}\right)^{i+1}}{\frac{1}{a} + \frac{1}{\kappa_1}} = \frac{p^{i+1} + (-1)^i q^{i+1}}{m^i q^i}.$$

If now the above values of the μ_{ij} and c_i are inserted in the expressions (1), the following final formulas are obtained.

$$E(z) = (N-1)\left\{1 + \frac{p}{Nm} + \left(\frac{p}{Nm}\right)^2 + \left(\frac{p}{Nm}\right)^3 + \cdots\right\}$$

$$E(z^2) = (N-1)^2\left\{1 + \frac{2}{N-1} - \frac{2(1-6pq)}{(N-1)Nmq} + \frac{pq(2-11pq)}{(Nmq)^2} - \frac{2(1+2pq-25p^2q^2)}{(N-1)(Nmq)^2} + \frac{2pq(1+3pq-30p^2q^2)}{(Nmq)^3} + \cdots\right\}$$

$$E(z^3) = (N-1)^3\left\{1 + \frac{6}{N-1} - \frac{3pq}{Nmq} + \frac{8}{(N-1)^2} - \frac{6(1-3pq)}{(N-1)Nmq} + \frac{2pq(1-5pq)}{(Nmq)^2} + \frac{4(1-4pq)(N-2)}{(N-1)^2Nmq} - \frac{24(1-5pq)}{(N-1)^2Nmq} - \frac{6(1-11pq+40p^2q^2)}{(N-1)(Nmq)^2} + \frac{6pq(1-16pq+55p^2q^2)}{(Nmq)^3} + \frac{60pq(1-4pq)(N-2)}{(N-1)^2(Nmq)^2} + \cdots\right\} \tag{3}$$

$$E(z^4) = (N-1)^4\left\{1 + \frac{12}{N-1} - \frac{8pq}{Nmq} + \frac{44}{(N-1)^2} - \frac{12(1+2pq)}{(N-1)Nmq} - \frac{2pq(2-21pq)}{(Nmq)^2} + \frac{16(1-4pq)(N-2)}{(N-1)^2Nmq} + \frac{48}{(N-1)^3} - \frac{8(15-46pq)}{(N-1)^2Nmq} - \frac{12(3-44pq+138p^2q^2)}{(N-1)(Nmq)^2} + \frac{64pq(1-4pq)(N-2)}{(N-1)^2(Nmq)^2} + \frac{96(1-4pq)(N-2)}{(N-1)^3Nmq} + \frac{8(1-12pq+36p^2q^2)(4N^2-9N+6)}{(N-1)^3(Nmq)^2} + \frac{4pq(1-43pq+168p^2q^2)}{(Nmq)^3} + \cdots\right\}.$$

By considering the formation of terms, it can also be shown that the above expressions are correct to terms of order m^3, m^2, m^1, and m^0, respectively, in the parameter m. If m is large these expressions are considerably more accurate than the order N^{-3} would indicate since the lowest order terms neglected in these expressions are respectively N^4m^4, N^4m^3, N^4m^2, and N^4m.

3. Applications. To compare these moments with those of the chi-square distribution, consider the ratios of corresponding moments, both for the Poisson distribution and for the binomial distribution in the special case of $p = \frac{1}{3}$.

For the Poisson distribution, these ratios are

$$R_1 = 1$$

$$R_2 = 1 - \frac{1}{Nm} - \frac{1}{(Nm)^2}$$

$$R_3 = 1 + \frac{1}{2m} - \frac{4}{Nm}$$

$$R_4 = 1 + \frac{2}{N+3}\left\{\frac{3}{m} + \frac{1}{3m^2} - \frac{7}{Nm}\right\}.$$

For the binomial distribution with $p = \frac{1}{3}$, these ratios are

$$R_1 = 1 + \frac{1}{Nn} + \frac{1}{(Nn)^2} + \frac{1}{(Nn)^3}$$

$$R_2 = \left(1 - \frac{1}{n}\right)\left(1 + \frac{5}{2Nn} - \frac{7}{4N^2n^2}\right) - \frac{7}{4(Nn)^3}$$

$$R_3 = \left(1 - \frac{1}{n}\right)\left(1 - \frac{7}{4n}\right) + \frac{1}{Nn}\left(4 - \frac{13}{2n} + \frac{5}{2n^2}\right) + \frac{1}{N^2n^2}\left(1 - \frac{5}{n}\right) + \frac{5}{2(Nn)^3}$$

$$R_4 = 1 + \frac{N}{N+3}\left\{\frac{-2}{n} + \frac{1}{n^2} + \frac{1}{Nn}\left(-13 + \frac{37}{2n} - \frac{17}{2n^2}\right)\right.$$
$$\left. + \frac{1}{N^2n}\left(11 - \frac{67}{2n} + \frac{51}{2n^2}\right) + \frac{1}{N^3n^2}\left(\frac{31}{2} - \frac{51}{2n}\right) + \frac{17}{2N^4n^3}\right\}.$$

From these expressions the following table is constructed.

m	n	N	R_1	R_2	R_3	R_4
25	∞	3	1	.99	.97	1.01
25	75	3	1	1	.98	.97
5	∞	5	1	.96	.94	1.08
5	15	5	1.01	.96	.87	.84
2	∞	∞	1	1	1.25	1
2	∞	10	1	.95	1.05	1.19
2	∞	5	1	.89	.85	1.21
2	6	∞	1	.83	.59	.69
2	6	10	1.02	.87	.64	.64
2	6	5	1.03	.90	.69	.62
1	∞	25	1	.96	1.34	1.22
1	∞	10	1	.89	1.10	1.39
1	∞	5	1	.76	.70	1.44
1	3	25	1.01	.69	.31	.41
1	3	10	1.03	.72	.35	.38
1	3	5	1.07	.77	.41	.36

For $m \geq 5$ these ratios are close to unity even for N as small as 5; hence it appears that the chi-square approximation is satisfactory as long as $m \geq 5$.

For $m \leq 2$ most of these ratios differ considerably from unity, particularly for the binomial distribution. Since R_1 is practically constant, the reduction in R_2 here indicates that the chi-square approximation will contain too many extreme values. For the Poisson distribution there is an increase in R_4 to compensate slightly for this decrease in R_2 so that the 5 percent points, for example, would not differ very much. The use of the chi-square approximation would therefore tend to give slightly too few significant results when they exist. For the binomial distribution, however, there is a decrease in both R_3 and R_4, so that the distribution tends toward normality; consequently the chi-square approximation will contain far too many extreme values and the 5 percent point will be much too large. This situation becomes slightly worse with increasing N.

4. Conclusions. From a consideration of the approximations for the first four moments of the distribution of the index of dispersion, it appears that the chi-square approximation is highly satisfactory provided that $m \geq 5$. For smaller values of m, the approximation is still fairly accurate for the Poisson distribution but not for the binomial distribution. For decreasing small values of m there is an increasing tendency to claim compatibility between data and theory when it does not exist; hence the binomial index must be handled carefully in such situations. These general conclusions are in agreement with the specialized results of Cochran and Sukhatme.

The semi-invariant technique for problems such as this is exceedingly laborious and is of questionable accuracy. The coefficients in Fisher's heavier formulas are so large that increased accuracy comes slowly with increased accuracy of order of terms. In addition, there are numerous typographical mistakes in Fisher's formulas, some of which are not easily detected. The formulas (3) may be used to investigate the accuracy of the chi-square approximation for situations not covered in the numerical table, but they are of questionable accuracy, when m is small, for N as small as 5.

REFERENCES

[1] P. V. Sukhatme, "On the distribution of chi-square in samples of the Poisson series," *Jour. Roy. Stat. Soc.*, Vol. 101 (1938), pp. 75–79.

[2] W. G. Cochran, "The chi-square distribution for the binomial and Poisson series with small expectations," *Annals of Eugenics*, Vol. 7 (1936), pp. 207–17.

[3] R. A. Fisher, "Moments and product moments of sampling distributions," *Proc. London Math. Soc.*, Series 2, Vol. 30 (1930), pp. 199–238.

ON SERIAL NUMBERS

By E. J. Gumbel

New School for Social Research

In this paper we consider a continuous variate and unclassified observations. It is well known that there are two step functions, which we may trace for a given series of observations. We will show that the differences between the two ways of plotting play an important rôle for certain graphical methods used by engineers.

To obtain one and only one series of observations we adjust the cumulative frequencies. The corrections thus introduced depend upon the theoretical distribution which is adequate for the observations. Later we deal with the relation between serial numbers and grades. Finally we construct confidence bands for the comparison between theory and observations.

1. Theory and observations. If we arrange n observations in order of increasing magnitude, and write each as often as it occurs, there will be a first, x_1, the smallest value, a second, x_2, an mth, x_m the penultimate, x_{n-1}, and the last, x_n, i.e., the greatest value. The index m is called the observed cumulative frequency, or simply the rank. It is usual to draw the observations x_m along the abscissa, and the rank m along the ordinate. The step function starts with a vertical line from the value x_1 of the abscissa to the point with the coordinates $1, x_1$, and in general consists of the horizontal lines from the point m, x_m to the point m, x_{m+1} and the vertical lines from the point m, x_{m+1} to the point $m + 1, x_{m+1}$. The step function ends with the point n, x_n. We call this graph the step function (m, x_m). However, another step function which is derived from the observations arranged in decreasing magnitude is equally legitimate. This step function starts from the point with the coordinates $0, x_1$, and in general consists of the horizontal lines from the point $m - 1, x_m$ to $m - 1, x_{m+1}$ and the vertical lines from the point $m - 1, x_{m+1}$ to the point m, x_{m+1} and ends with the point $n - 1, x_n$. We call it the step function $(m - 1, x_m)$. Let $F(x)$ be the probability of a value equal to or less than x. Then the continuous theoretical curve, the ogive, which we compare to the step functions is $nF(x), x$. The question is whether we have to use the step function (m, x_m) or the step function $(m - 1, x_m)$.

The differences between the two ways of plotting are rarely mentioned in the statistical literature. If we plot instead of the rank m the reduced rank m/n, the differences between the two ways of plotting are of the order $1/n$. It is generally tacitly assumed that this difference may be neglected. This will not hold if n is small.

In the following we show two other ways of plotting the observations where the differences between the two observed curves play an important role. Sup-

pose that the probability $F(x)$ and the density of probability, $f(x)$, henceforth called the distribution are such that it is possible to introduce a reduced variate

$$z = \frac{x - a}{b}, \tag{1}$$

which has no dimension. In general, the constant a will be a certain mean, and the constant b a certain measure of dispersion. Furthermore, the constants may be linear functions of these characteristics. Neither the probability $G(z)$ of a value equal to or less than z

$$G(z) = F(x), \tag{2}$$

nor the reduced distribution

$$g(z) = bf(a + bz) \tag{3}$$

contain constants. The *equiprobability test* consists in the following procedure: We attribute to the mth observation x_m the relative frequency m/n, and determine from a probability table a value z, such that

$$G(z) = m/n. \tag{4}$$

The variate x is plotted on the ordinate, and the reduced variate z on the abscissa. Then the points x_m , z must be situated close to the straight line (1). To apply this comparison between theory and observations, we need not even calculate the constants. For the normal distribution the application of this test is greatly facilitated by the use of probability paper.

The difficulty is that we may as well choose the frequency

$$G(z) = (m - 1)/n, \tag{4'}$$

and determine the corresponding values of z. Therefore, we have two lines (1) instead of one. The difference between the two series will be large for the first and last few observations. For the first series the last observation cannot be plotted on probability paper; for the second series the first observation cannot be plotted.

The same difficulty exists for the "return period." If the observations of a continuous variate are made at regular intervals in time which are taken as units, we may as in [4] define the theoretical return periods $T(x)$ of a value equal to or greater than x as

$$T(x) = \frac{1}{1 - F(x)}. \tag{5}$$

The comparison of the theoretical with the observed return periods gives a test for the validity of a theory. However, there are two series of observations, namely, the exceedance intervals

(6) $$'T(x_m) = \frac{n}{n-m}; \qquad m = 1, 2 \cdots n-1$$

and the recurrence intervals

(7) $$''T(x_m) = \frac{n}{n-m+1}; \qquad m = 1, 2 \cdots n.$$

The two expressions (6) and (7) differ widely for the high ranks. The penultimate observation, for example, has an exceedance interval n, whereas the recurrence interval is only $n/2$. This contradiction and the difficulty arising for the equiprobability test show that the question of choosing the observed cumulative frequency of the mth observation has a practical significance.

The equiprobability test and the comparison between the observed and the theoretical return period may be combined on probability paper. The variate x is plotted on the ordinate, the reduced variate y on the abscissa. But instead of y we write the probability $F(x)$ and the return period $T(x)$. If the theory holds, the observations must be scattered around the straight line (1).

But all these methods presuppose that we know whether we have to attribute to x_m the rank m or the rank $m-1$. Sometimes a compromise has been proposed which consists in attributing to x_m neither m nor $m-1$, but the arithmetic mean of both, $m - \frac{1}{2}$. In other words, the index m is no longer considered to be an integer. In such cases, we call *m the serial number.*

The corrected frequency $m - \frac{1}{2}$ may be accepted for the comparison between the step function and the probability curve. However, for the return period and for the equiprobability test this method leads to serious difficulties. The corrected return periods, which have been proposed by Hazen [7] and have been used by M. Kimball [8] are

(6) $$T(x_m) = \frac{n}{n-m+1/2}.$$

The last among n observations has a return period $2n$. This idea does not seem to be sound. No statistical device can increase the number of observations beyond n.

2. The adjusted frequency of the *m*th observation. The use of m, $m-1$, or $m - \frac{1}{2}$ as frequency of the mth observations amounts to considering the mth value as being fixed. To obtain one and only one step function we consider x_m as a statistical variate. This will lead to the determination of the most probable serial number and of the corresponding probability as a function of m and n.

The mth observation is such that there are $m-1$ observations below it and $n-m$ observations above it. Consequently, the distribution $w_n(x, m)$ of the mth observation is

(9) $$w_n(x, m) = \binom{n}{m} m[F(x)]^{m-1}[1 - F(x)]^{n-m} f(x).$$

The variate x_m is simply called x as each value of x has a certain density of probability of being the mth. To distinguish between (x) and $w_n(x,m)$, the first distribution is referred to as the *initial* distribution. For some simple initial distributions it is possible to calculate exactly the mean and the standard error of the distribution (9). This has been done by Karl Pearson [10] for the normal, the uniform, the exponential, and other skew distributions. The results are very complicated, and do not allow any immediate practical applications.

In the following we determine therefore instead of the mean the mode of the mth value. The most probable mth value for which we write $\tilde{x}_m$ is the solution of

$$\frac{d \log w_n(x,m)}{dx} = 0.$$

We obtain from (9)

$$\frac{m-1}{F(\tilde{x}_m)} f(\tilde{x}_m) - \frac{n-m}{1-F(\tilde{x}_m)} f(\tilde{x}_m) = -\frac{f'(\tilde{x}_m)}{f(\tilde{x}_m)}. \tag{10}$$

In this equation m is counted in order of increasing magnitude. If we choose the inverse order we obtain the same equation, if we replace the index m by $n - m + 1$. Therefore the following results are independent of the order of counting m.

Equation (10) gives the most probable value $\tilde{x}_m$ as a function of m and n. The function depends upon the distribution.

A rough, first trial solution of (10) may be obtained if we confine our interest to values where neither m nor $n - m$ is small in comparison to n, that is, values which are not extreme. Suppose m to be of the order $n/2$. For increasing numbers of observations, the expression on the left side of (10) become large compared to the right side provided the derivative remains finite, as is generally the case. If we neglect the right-hand member, $\tilde{x}_m$ is the solution of

$$F(\tilde{x}_m) = \frac{m-1}{n-1}. \tag{11}$$

This expression holds for the uniform distribution where $f'(x) = 0$.

The following exact solution is valid for any number of observations and any serial number. Equation (10) will be used in two different ways: First, we suppose m to be known, we determine the probability $F(\tilde{x}_m)$ of the most probable mth value as a function of m and n, and attribute this probability to the mth observation x_m. By doing so, the probability of the most probable mth value becomes the *adjusted frequency* of the mth observation. This leads to one and only one series of observations, and settles our initial question. Later, in section 3, we suppose $F(\tilde{x}_m)$ to be known, and compute the corresponding most probable mth observation. This leads to an estimate of the grades (or partition values) from the serial numbers.

To obtain $F(\tilde{x}_m)$ from (10) we introduce an expression $\sigma^2(x_m)$ by stating

$$[\sigma^2(x_m)n] = F(\tilde{x}_m)[1 - F(x_m)]f^{-2}(\tilde{x}_m) . \tag{12}$$

The brackets are meant to indicate that the product on the left side does not depend upon n. We shall show later that $\sigma^2(x_m)$ is under certain conditions the variance of the mth observation. For the present purpose however this significance is not required. Multiplication of (10) by (12) leads to

$$m - 1 + F(\tilde{x}_m) - nF(\tilde{x}_m) = -f'(\tilde{x}_m)[\sigma^2(x_m)n], \tag{13}$$

or

$$F(\tilde{x}_m) = \frac{m-1}{n-1} + \frac{f'(\tilde{x}_m)[\sigma^2(x_m)n]}{n-1} . \tag{14}$$

The adjusted frequency in (14) is similar to (11). Another expression for the adjusted frequency, derived from (13), is

$$F(\tilde{x}_m) = \frac{m - \frac{1}{2}}{n} + \frac{1}{n}\,(F(\tilde{x}_m) - \tfrac{1}{2} + f'(\tilde{x}_m)[\sigma^2(x_m)n]). \tag{15}$$

The adjusted frequency is the compromise $\frac{m - \frac{1}{2}}{n}$ plus an expression

$$\frac{D}{n} = \frac{1}{n}\,(F(\tilde{x}_m) - \tfrac{1}{2} + f'(\tilde{x}_m)[\sigma^2(x_m)n]). \tag{16}$$

The correction, D, defined by (16) depends upon the initial distribution and has no dimension. In general, it will depend upon the constants which exist in the distribution. If the distribution $f(x)$ may be written in a reduced form (3), the correction[1]

$$D = G(z) - \tfrac{1}{2} + g'(z)[\sigma^2(z)n] \tag{17}$$

depends only upon the dimensionless reduced variate z. For a given initial distribution we choose numerical values for the probability $G(z) = F(\tilde{x}_m)$ calculate $g'(z)$ and

$$[\sigma^2(z)n] = \frac{G(z)(1 - G(z))}{g^2(z)} . \tag{18}$$

From (16) we compute a table for the corrections D as a function of the adjusted frequencies $F(\tilde{x}_m)$ and obtain for given n the serial number m as a function of the adjusted frequencies by

$$m = nF(\tilde{x}_m) + \tfrac{1}{2} - D. \tag{19}$$

These serial numbers will not be integers. The adjusted frequency $F(\tilde{x}_m)$ for the mth observation will then be obtained by linear interpolation.

[1] In previous articles [3, 6] we started from another interpretation of the corrected frequencies and obtained slightly different corrections.

The value and the sign of the correction D depends upon the distribution. For the asymmetrical exponential distribution, for example, the correction

$$D = -\tfrac{1}{2}, \tag{19'}$$

is independent of the variate. This means that we have to use exclusively the step function $(m - 1, x_m)$ as being the best way of plotting. The observed adjusted return periods are the recurrence intervals.

For a symmetrical reduced distribution we have

$$1 - G(-z) = G(z); \qquad g(-z) = g(z); \qquad g'(-z) = -g'(z). \tag{20}$$

Therefore, the reduced correction will be

$$D(-z) = -D(z). \tag{21}$$

For the two reduced values z and $-z$ of a symmetrical variate the corrections have the same size, but different signs.

A relation similar to (21) holds for two asymmetrical reduced distributions $g_1(z)$ and ${}_1g(z)$, which are symmetrical one to another in the sense

$$G_1(z) = 1 - {}_1G(-z); \qquad g_1(z) = {}_1g(-z); \qquad g_1'(z) = -{}_1g'(-z). \tag{22}$$

Then, the corrections are

$$D_1(-z) = -{}_1D(z). \tag{23}$$

For any initial distribution $f(x)$ we read from (19) the adjusted frequency

$$F(\tilde{x}_m) = \frac{m - \tfrac{1}{2} + D}{n}, \tag{24}$$

even for a small number of observations. The question whether to choose m/n or $(m - 1)/n$ as observed cumulative frequency is settled by (24). We obtain one observed step function, one series for the equiprobability test, and one series of observed return periods

$$T(\tilde{x}_m) = \frac{n}{n - m + \tfrac{1}{2} - D}, \tag{25}$$

which have to be compared to the theoretical continuous curves.

3. Estimates for the grades. In the following we use the fundamental formula (15) to determine interesting grades through the mth values.

We use the term *grade* for the value of a statistical variate which corresponds to a given cumulative probability $F(x)$ say, $F(x) = \frac{1}{4}; \frac{1}{2}; \frac{3}{4}$ for quartiles; $F(x) = \frac{1}{10}, \cdots \frac{9}{10}$ for deciles, and so on. For a given grade, the probability $F(x)$ the density of probability $f(x)$ and its derivative are known, and m is unknown. The value of m obtained from (15), henceforth called the most probable serial number $\tilde{m}$, is the solution of

$$\tilde{m} = nF(x) + 1 - F(x) - f'(x)F(x)(1 - F(x))f^{-2}(x). \tag{26}$$

The corresponding "observed" value $x_{\tilde{m}}$ is obtained by interpolation between two observed values x_{m-1} and x_m, such that

$$m - 1 < \tilde{m} < m.$$

For the median, x_0, the most probable serial number $\tilde{m}_0$ is

$$\tilde{m}_0 = \frac{n+1}{2} - \frac{f'(x_0)}{4f^2(x_0)}. \tag{27}$$

The median x_0 itself enters into (27). It has to be eliminated through the condition $F(x_0) = \frac{1}{2}$. For the exponential distribution for example we find

$$\tilde{m}_0 = \frac{n}{2} + 1. \tag{27'}$$

The most probable serial number of the median for a symmetrical distribution is

$$\tilde{m}_0 = \tfrac{1}{2}(n+1). \tag{28}$$

This is the usual estimate of the median for any distribution. The estimate obtained from (27) is smaller (larger) than the usual estimate if the median is smaller (larger) than the mode. The difference between the two estimates is due to the fact, that (27) makes use of information about the theoretical distribution whereas this information (if available) is neglected by the usual method.

For symmetrical distributions the most probable serial numbers $\tilde{m}_1$ and $\tilde{m}_2$ for two symmetrical grades defined by F_1 and $F_2 = 1 - F_1$ are according to (26) related by

$$\begin{aligned} \tilde{m}_1 &= nF_1 + 1 - (F_1 + f_1'F_1(1 - F_1)f_1^{-2}) \\ \tilde{m}_2 &= n(1 - F_1) + (F_1 + f_1'F_1(1 - F_1)f_1^{-2}). \end{aligned} \tag{29}$$

The members in brackets have the same size, but opposite signs. Another expression for $\tilde{m}_2$ is

$$\tilde{m}_2 = (n+1) - [nF_1 + 1 - F_1 - f_1'F_1(1 - F_1)f_1^{-2}]$$

so that, for symmetrical distributions

$$\tilde{m}_1 + \tilde{m}_2 = n + 1. \tag{30}$$

This is to be expected as the mth value counted upwards is the $(n - m + 1)$st value counted downwards.

For the two quartiles q_1 and q_2 the most probable serial numbers $\tilde{m}(q_1)$ and $\tilde{m}(q_2)$, obtained from (29) are

$$\tilde{m}(q_1) = \frac{n+3}{4} - \frac{3}{16}\frac{f'(q_1)}{f^2(q_1)}; \qquad \tilde{m}(q_2) = \frac{3n+1}{4} - \frac{3}{16}\frac{f'(q_2)}{f^2(q_2)}, \tag{31}$$

where q_1 and q_2 have to be eliminated by the use of

$$F(q_1) = \tfrac{1}{4};\ F(q_2) = \tfrac{3}{4}.$$

For the uniform, the normal and the exponential distribution we obtain the two quartiles from

$$\tilde{m}(q_1) = \frac{n+3}{4} \quad ; \qquad \tilde{m}(q_2) = \frac{3n+1}{4}$$

$$\tilde{m}(q_1) = \frac{n}{4} + .352; \qquad \tilde{m}(q_2) = \frac{3n}{4} + .648 \tag{31'}$$

$$\tilde{m}(q_1) = \frac{n}{4} + 1 \quad ; \qquad \tilde{m}(q_2) = \frac{3n}{4} + 1$$

respectively. The last result may also be found from (19') and (24). These estimates differ from the usual estimates by the reason given above.

We now apply the notion of a grade to certain characteristics which are *otherwise* defined. A certain characteristic, say, the mode $\tilde{x}$ or the mean $\bar{x}$ have for a given distribution the probabilities $F(\tilde{x})$ or $F(\bar{x})$ respectively. These probabilities may be used to define a grade. We determine the corresponding mth value from (26), and obtain an estimate of the mode or the mean, interpreted as grades by interpolation between the observed mth values. For a symmetrical distribution these estimates of the mode and mean are identical with the estimates of the median. For an asymmetrical distribution, the most probable serial number $\tilde{m}(\tilde{x})$ of the mode becomes according to (26)

$$\tilde{m}(\tilde{x}) = (n-1)F(\tilde{x}) + 1. \tag{32}$$

Usually, the mode $\tilde{x}$ of a continuous variate is estimated by another procedure. The observations are arranged in certain cells. One of them has the largest relative frequency. It will contain the mode. To find its position within the cell, an interpolation formula is applied which reproduces the content of this cell and of the two adjacent cells. By choosing different lengths for the cells and different origins for the classification, the mode can be shifted to the right or to the left. Formula (32) furnishes a determination of the mode from the observations according to the theory, such that the arrangement of the observations into different cells is not needed. Of course, this method can be applied only if we know the theoretical distribution $f(x)$. The problem how to estimate the mode is important for distributions where one of the constants may be interpreted as the mode or as a function of the mode [1, 4].

4. Standard errors of the estimates. The numerical work involved in the method (26) of estimating the grades is very small. To obtain the standard errors of these estimates we consider the asymptotic properties of the distribution (9). The following results hold therefore only for large numbers of observation. Besides we assume, that the serial number m is of the order $n/2$, i.e. not extreme. It has been shown [2] that under these conditions the distribution

of the mth value converges toward a normal distribution with a standard error $\sigma(x_m)$, where

$$\sigma(x_m)\sqrt{n} = \frac{1}{f(x)}\sqrt{F(x)(1 - F(x))}. \tag{33}$$

Although this standard error does not contain m explicitly, it has a clear meaning for any value of x as we know from (26), which observation we have to attribute to the probability $F(x)$. The classical proof about the approximate normality of the distribution of the median in large samples is a special case of this convergence and the classical standard error of the median,

$$\sigma(x_0)\sqrt{n} = \frac{1}{2f(x_0)}, \tag{34}$$

is a special case of (33). The square root in (33) is maximum for $F(x) = \frac{1}{2}$. Therefore,

$$\sigma(x_m)\sqrt{n} \leqq \frac{1}{2f(x)}. \tag{35}$$

If the variate x may be reduced through the linear transformation (1) the standard error $\sigma(z)$ of the reduced variate, called reduced standard error

$$[\sigma(z)\sqrt{n}] = \frac{1}{g(z)}\sqrt{G(z)(1 - G(z))}, \tag{36}$$

may be calculated as a function of z where z corresponds to x_m. To call attention to the fact that these numerical values do not depend upon n, they are written in brackets. The standard error of the estimate for x_m is, according to (2) and (3)

$$\sigma(x_m) = \frac{b}{\sqrt{n}}[\sigma(z)\sqrt{n}]. \tag{37}$$

Since the constant b is a measure of dispersion, the standard error of the estimate of the mth value is proportional to the standard deviation of the variate. The factors b and $1/\sqrt{n}$ show that the standard error of the mth value is of the same structure as the standard error of the arithmetic mean.

For symmetrical distributions the standard error (33) of the mth value is also a symmetrical function. The standard errors of the estimate of the two quartiles, and generally of the estimates of two grades defined by F and $1 - F$, are then identical. If the mode coincides with the median, the corresponding standard error of the mth value is a minimum. For a symmetrical U-shaped distribution, however, where the density of probability passes through a minimum at the center of symmetry, the median has the largest standard error among the mth values. An example for such a distribution has been given by Leavens [9]. As the distribution of the mth value converges towards a normal distribution, it is legitimate to attribute to the mode of the mth value the standard error (33).

Therefore, for a large number of observations (33) gives the standard error of our estimate of the grades. The standard errors of the estimates (31) of the quartiles are

$$\sigma(q_1)\sqrt{n} = \frac{\sqrt{3}}{4f(q_1)}; \qquad \sigma(q_2)\sqrt{n} = \frac{\sqrt{3}}{4f(q_2)}. \tag{38}$$

The arithmetic mean in its usual definition is not an mth value. Its standard error $\sigma(\bar{x})$, where

$$\sigma(\bar{x})\sqrt{n} = \sigma, \tag{39}$$

will, therefore, fall outside of the range of the standard errors of the mth values. (See graph 1.) If the distribution $f(x)$ is such that the standard deviation does not exist, it is legitimate to estimate the arithmetic mean as a grade, and calculate it from the corresponding most probable mth value by introducing $F(\bar{x})$, $f(\bar{x})$ and $f'(\bar{x})$ into (26). The standard error of the arithmetic mean interpreted as a grade is

$$\sigma(\bar{x})\sqrt{n} = \frac{1}{f(\bar{x})}\sqrt{F(\bar{x})(1 - F(\bar{x}))}. \tag{40}$$

If we use this estimate of the arithmetic mean for distributions where σ exists, the usual determination of the mean will be more (less) precise than its estimate as a grade if

$$\sigma f(\bar{x}) \lesseqgtr \sqrt{F(\bar{x})(1 - F(\bar{x}))}. \tag{41}$$

The standard error of the mode estimated as a grade is

$$\sigma(\tilde{x})\sqrt{n} = \frac{1}{f(\tilde{x})}\sqrt{F(\tilde{x})(1 - F(\tilde{x}))}. \tag{42}$$

As the standard error of any characteristic depends upon the way it is estimated from the observations, the standard errors of the mode or mean interpreted as grades differ from the usual standard errors.

5. The most precise grade. Equation (33) may be used to define a new grade which has interesting properties. The standard error (33) of the estimate of the mth value is a function of F. We ask whether it possesses a minimum (maximum). The corresponding value of the variate, $\hat{x}$, may be called *the most (least) precise* m*th value* or the most (least) precise grade. To obtain $\frac{d\sigma(x_m)}{dF}$ it is sufficient to calculate from (33)

$$\frac{nd \log \sigma^2(x_m)}{dx} = \frac{2n\sigma'(x_m)}{\sigma(x_m)}.$$

Therefore the most (least) precise grade is the solution of

$$\frac{f(x)}{F(x)} - \frac{f(x)}{1 - F(x)} - \frac{2f'(x)}{f(x)} = 0. \tag{43}$$

This expression does not vanish if either $F(x) = \frac{1}{2}$ or $f'(x) = 0$. It vanishes if both conditions hold simultaneously. For a symmetrical distribution passing through a mode (minimum), the mode (minimum), estimated as a grade, is the most (least) precise grade. Equation (43) may be written

$$f'(x)f^{-2}(x)F(x)(1 - F(x)) = \tfrac{1}{2} - F(x).$$

If we introduce this expression into (16), we obtain $D = 0$ and

$$F(\hat{x}) = \frac{m - \frac{1}{2}}{n}. \tag{44}$$

The most precise m*th value is such that the adjusted frequency is the arithmetic mean of the frequencies* m/n *and* (m − 1)/n.

The most precise mth value $\hat{x}$ cannot be calculated from the observations alone. It may be estimated in the same way as any grade by introducing the values $F(\hat{x})$, $f(\hat{x})$ and $f'(\hat{x})$ into equation (26).

To show the difference between the most precise grade and the mode we apply the procedure developed above to a skew distribution. The reduced distribution of the largest value $g(y)$ and the probability $G(y)$ are

$$g(y) = e^{-y}G(y); \qquad G(y) = e^{-e^{-y}}. \tag{45}$$

The relation (1) between the reduced variate for which we write y instead of z and the largest value x is

$$x = u + \frac{y}{\alpha}. \tag{46}$$

where $u = \tilde{x}$ is the mode and

$$\frac{1}{\alpha} = \frac{\sqrt{6}}{\pi}\sigma. \tag{47}$$

The most probable serial number $\tilde{m}(u)$ of the mode, obtained from (32) is

$$\tilde{m}(u) = \frac{n + e - 1}{e}. \tag{48}$$

This equation may be used for an estimate of the constant u.

The reduced variance $\sigma^2(y)$ obtained from (36) and (45) is

$$(\sigma^2(y)\sqrt{n}) = e^{2y}(e^{e^{-y}} - 1). \tag{49}$$

A table for the reduced standard error $\sigma(y)\sqrt{n}$ has been given in a previous publication [6]. The value $\sigma(y)\sqrt{n}$ is plotted in figure 1 for probabilities $G(y)$ from 0.01 to 0.95. The standard error has a minimum for a value of y located to the left of the mode $\tilde{y} = 0$. On the same graph are plotted the reduced standard errors for the normal distribution. As the normal reduced variate z differs from the reduced variate y, two different scales are used for the variates. The standard error of the estimate (48) of the mode interpreted as a grade, obtained by introducing $y = 0$ into (49) is

$$\sigma(u)\sqrt{n} = \frac{1}{\alpha}\sqrt{e-1} = 1.02205\sigma. \tag{49'}$$

The most precise grade is

$$\hat{x} = u + \frac{\hat{y}}{\alpha},$$

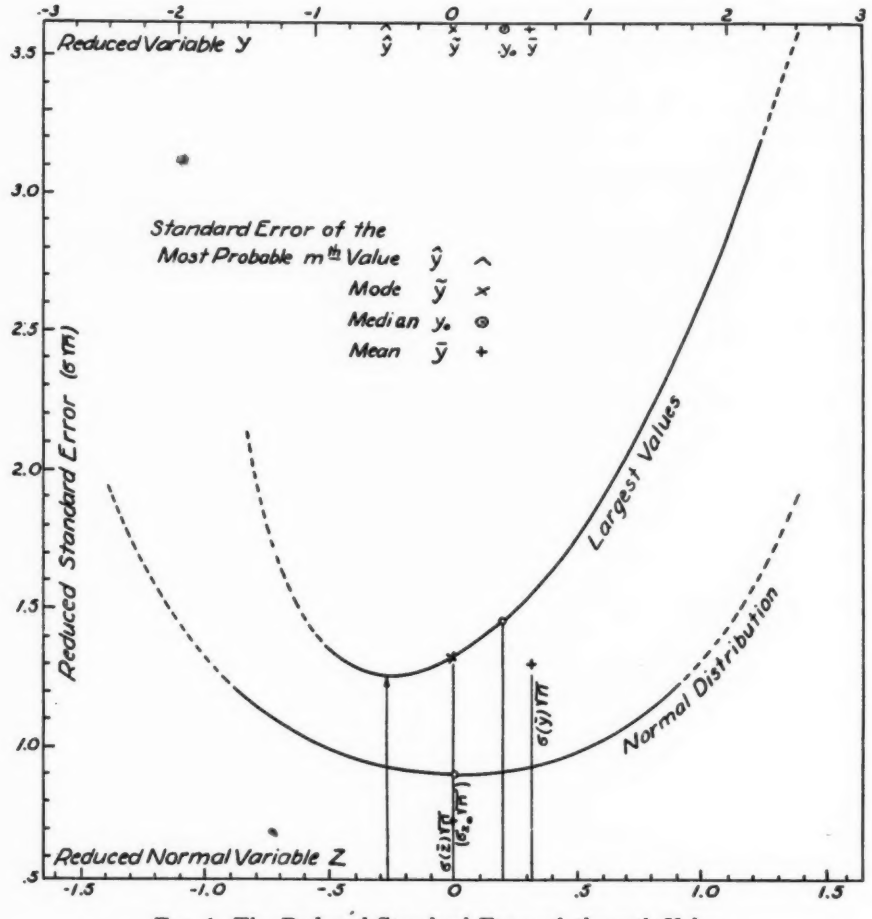

FIG. 1. The Reduced Standard Error of the mth Value

where $\hat{y}$ is the value of the reduced variate, for which the standard error (49) is minimum. We obtain from (49) and (45) the numerical values

$$\hat{y} = -.46601;\ G(\hat{y}) = .20319;\ \sigma(\hat{x})\sqrt{n} = .96887\sigma. \tag{50}$$

The standard error of the most precise grade is 3 per cent smaller; the standard error of the mode, estimated as a grade, is 2 per cent larger than the standard error of the mean.

6. Confidence bands. The standard errors (33) of the grades may be used in a general way for the construction of *confidence bands* obtained from curves which control the fit between theory and observation. Consider first the observed stepfunction $(m - \frac{1}{2}, x_m)$ and the theoretical ogive $nF(x)$, x. The variate x is plotted along the abscissa, the cumulative frequency along the ordinate. Now, for large n any theoretical value x, which is not extreme, may be interpreted as an mth value having a normal distribution and a standard error $\sigma(x_m)$. At each point of the graph of $nF(x)$, x which is not extreme, we construct a segment of length $2\sigma(x_m)$ parallel to the x axis, the midpoint of the segment being on the theoretical ogive. In other words, we add the standard error $\sigma(x_m)$ to, and subtract it from, any corresponding value x, and attribute $nF(x)$ to the beginning and end of these intervals. By this procedure we obtain two curves $nF(x)$, $x \mp \sigma(x_m)$. For each observation there exists a probability $P = .68268$ that it will be contained within the interval $x \mp \sigma(x_m)$.

If we apply another hypothesis to the same observations, or choose other values for the constants, we reach, of course, other control curves. Of two competing hypotheses the one is to be preferred where the band contains a larger number of observations.

The same method may be applied to the equiprobability test and to the comparison of the observed and theoretical return periods [6]. This procedure is legitimate for all values which are not extreme.

In the following, we construct the confidence bands for the normal distribution

$$g(z) = \frac{1}{\sqrt{\pi}} e^{-z^2}. \tag{51}$$

The variate x is related to the reduced variate z by (1), which, in this case, becomes

$$x = \bar{x} + \sigma\sqrt{2}z. \tag{52}$$

The probability $G(z)$ is

$$G(z) = \tfrac{1}{2}(1 + \Phi(z)), \tag{53}$$

where $\Phi(z)$ stands for the Gaussian integral

$$\Phi(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, dt. \tag{54}$$

Formulae (36) and (53) lead to the reduced standard error

$$\sigma(z)\sqrt{n} = \frac{1}{2g(z)} \sqrt{1 - \Phi^2(z)}, \tag{55}$$

given in the table, col. 6. The standard errors $\sigma(x_m)$ of the mth values obtained from (37) (52) and (55) are

$$\sigma(x_m) = \frac{\sigma\sqrt{2}}{\sqrt{n}} [\sigma(z)\sqrt{n}]. \tag{56}$$

As a numerical example we choose the annual precipitations observed in 51 meteorological stations in Paris and its surroundings in the year 1938. We suppose that the differences between the 51 observations are only due to chance. The stepfunction $m - \frac{1}{2}$, x_m is plotted in figure 2. To obtain the theoretical ogive we compute the constants in (52). They are

$$\bar{x} = 571.92; \sigma\sqrt{2} = 38.52. \tag{57}$$

The theoretical values x obtained from (52), the cumulative frequencies $nF(x)$ obtained from the table of the Gaussian integral [11] and the standard errors

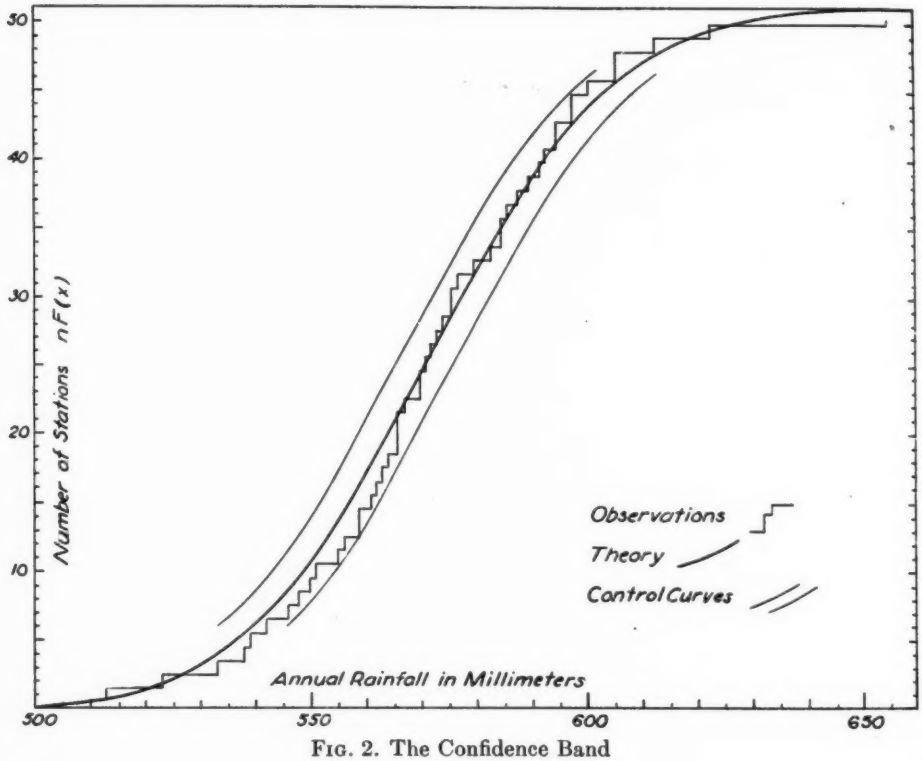

FIG. 2. The Confidence Band

$$\sigma(x_{m'}) = 5.393\ [\sigma(z)\sqrt{n}], \tag{58}$$

obtained from (56) are given in the columns 2 to 5 and 7 of Table I.

We trace in figure 2 the theoretical curve $nF(x)$, x and the confidence band obtained from col. 7. by the methods described above. All observations are contained within the control curves. We may accept the theory that the differences between the annual rainfalls observed in the 51 stations are only due to chance.

7. Conclusions. To test a statistical hypothesis for a continuous variate we use the ogive, the equiprobability method, based on (1), and the return periods

(5). The three tests may be combined on appropriate probability paper. As the rank of the mth observation x_m may be m or $m - 1$, we have two series of observations. To obtain one and only one series we use for the ogive the serial number $m - \frac{1}{2}$ provided that the number of observations is large. Generally, we attribute to x_m an adjusted frequency, namely, the probability (15) of the most probable mth value. The adjusted frequency is obtained from the serial number $m - \frac{1}{2}$ and a correction, D, equation (17), which depends upon the distribution. The correction is important for the three tests, and small n, furthermore, for the equiprobability test and the return periods for the extreme observations and any number n.

The same correction D is used for estimating a grade through its relation (26) to the corresponding most probable serial number $\tilde{m}$. For distributions, where the second moment does not exist, we estimate the arithmetic mean from a

TABLE I

Normal Confidence Band and Theoretical Frequencies of the Rainfalls

Reduced Variate $\pm z$ (1)	Variate		Frequency		Reduced Standard Error $\sigma(z)\sqrt{n}$ (6)	Standard Error $\sigma(x_m)$ (7)
	x (2)	x (3)	$51\,F(x)$ (4)	$51\,F(x)$ (5)		
0	571.91	571.9	25.50	25.50	.886	4.8
.2	564.2	579.6	19.82	31.18	.899	4.9
.4	556.5	587.3	14.58	36.42	.940	5.1
.6	548.8	595.0	10.10	40.90	1.012	5.5
.8	541.0	602.7	6.58	44.42	1.127	6.1
1.0	533.4	610.4	4.01	46.99	1.297	7.0
1.2	525.7	618.1	2.29	48.71		
1.4	418.0	625.9	1.22	49.78		
1.6	510.3	633.6	.60	50.40		
1.8	502.6	641.3	.28	50.72		

grade. For asymmetrical distributions we estimate the mode from a grade by (32) and (48).

In this case, we have to introduce a distinction between the mode and the most precise grade (43). The adjusted frequency and the estimates for grades may be used even for small numbers of observations.

The standard error of these estimates is obtained, equation (33) from the limiting, normal, form of the distribution of the mth value, which holds, provided the serial number is not extreme. To control a hypothesis we construct confidence bands, which are obtained from the standard errors of the grades.

REFERENCES

[1] R. A. Fisher and L. H. C. Tippett, "Limiting forms of the frequency distribution of the largest or smallest member of a sample," *Proc. Camb. Phil. Soc.*, Vol. 24, part 2 (1928), p. 180.

[2] E. J. GUMBEL, "Les valeurs extrèmes des distribution statistiques," *Annales de l'Institut Henri Poincaré*, Vol. 4 (1935), Paris, p. 115.
[3] E. J. GUMBEL, "Les valeurs de position d'une variable aléatoire," *Comptes Rendus*, Vol. 208, (1939), Paris, p. 149.
[4] E. J. GUMBEL, "The return period of flood flows," *Annals of Math. Stat.*, Vol. 12 (1942), p. 163.
[5] E. J. GUMBEL, "Simple tests for given hypotheses," *Biometrika*, Vol. 32 (1942), p. 317.
[6] E. J. GUMBEL, "Statistical control curves for flood discharge," *Trans. Am. Geoph. Union* (1942), Washington, p. 489.
[7] ALLEN HAZEN, *Flood Flows*, New York, John Wiley, 1930.
[8] B. F. KIMBALL, "Limited type of primary probability distribution applied to annual flood flows," *Annals of Math. Stat.*, Vol. 13 (1942), p. 318.
[9] DIXON H. LEAVENS, "Frequency distributions corresponding to time series," *Jour. Amer. Stat. Assoc.*, Vol. 26 (1931), p. 407.
[10] KARL AND MARGARET V. PEARSON, "On the mean character and variance of a ranked individual, and on the mean and variance of the interval between ranked individuals," *Biometrika*, Vol. 23, part 3, 4 (1931), p. 364; Vol. 24, part 1, 2 (1932), p. 203.
[11] *Tables of Probability Functions*, Federal Works Agency W.P.A. of New York City, 1941.

FITTING GENERAL GRAM-CHARLIER SERIES

By Paul A. Samuelson

Massachusetts Institute of Technology

1. Introduction. Since the last part of the nineteenth century at least, it has been common to represent a probability distribution by means of a linear sum of terms consisting of a parent function and its successive derivatives. Usually the parent function is the Type A or normal curve, as discussed by Gram [1], Bruns [2], Charlier [3], and numerous others. In addition there have been generalizations in various directions: for example, the Type B expansion in terms of the Poisson parent function and its successive finite differences.

Unlike these two types, which have a definite probability interpretation, another generalization involves the use of other parent functions and their derivatives (or differences) to give an approximate representation of a given frequency curve. With this process is associated the names of Charlier, Carver [4], Roa [5], and many others. Two general methods by which the equating of moments of the fitted curve and the given distribution yield the appropriate coefficients have been given by Charlier and Carver respectively. An account of the latter's technique is more accessible to the average English speaking statistician.

It is the purpose of the present discussion to indicate how the Charlier method may be simplified, and can be used to replace the Carver method. In doing so, I am following up the oral suggestion made some years ago by Professor E. B. Wilson of Harvard, that repeated integration by parts will yield the requisite coefficients very simply. At the same time certain methods implicit in the work of Dr. A. C. Aitken [6] show how the use of a moment generating function can often lighten the algebraic analysis. There will also be a brief indication of analogous results for general finite difference parent families; and attention will be called to a troublesome historical blunder which has permeated the statistical literature.

2. Alternative methods. Avoiding the overburdened expression generating function, I shall consider parent functions, called $f(x)$, with the restrictive properties:

a) Moments of all order of $f(x)$ exist.

b) Derivatives of any required order exist with appropriate continuity.

c) There exist high order contact at the extremities of the distribution as defined below.

Mathematically,

a) $\int_{-\infty}^{\infty} x^k f(x)dx$ is finite for all positive integral values of k

and

c) $\lim_{x\to\pm\infty} x^j f^k(x) = 0$ for all positive integral values of j and k.

These conditions suffice for many statistically interesting cases, but where desirable they can be lightened. Thus, derivatives may only be defined "almost everywhere," and there may be finite instead of infinite limits to the distribution, etc.

Given an arbitrary frequency curve $F(x)$, we shall suppose it to be formally expanded in the series

$$F(x) \sim a_0 f(x) + a_1 f'(x) + a_2 f''(x) + \cdots + a_n f^n(x) + \cdots . \tag{1}$$

For convenience in what follows, we shall assume that all distributions are given in terms of relative frequency so that the area under both f and F is equal to unity, so that a_0 may be taken as unity. The suppressed absolute frequency can clearly be restored at any time by multiplication of both sides with the appropriate constant. Also for algebraic convenience, many writers consider the slightly modified form of the expansion

$$F(x) \sim A_0 f(x) - \frac{A_1}{1!} f'(x) + \frac{A_2}{2!} f''(x) + \cdots \frac{(-1)^n A_n}{n!} f^n(x) + \cdots .$$

It is assumed without discussion that the first n coefficients in such a series are to be determined by equating the first n moments of each side.

I shall prove the two following identities:

$$(-1)^n a_n = L_n(F) - \sum_0^{n-1} L_{n-j}(f)(-1)^j a_j , \tag{2}$$

where

$$L_i(f^i) = \frac{\int_{-\infty}^{\infty} x^j f^i(x) dx}{j!} .$$

Alternatively

$$A_n = \sum_{i=0}^{n} \binom{n}{i} \frac{d^i}{d\alpha^i} \left(\frac{1}{\int_{-\infty}^{\infty} f e^{\alpha x} \, dx} \right)_{\alpha=0} \int_{-\infty}^{\infty} x^{n-i} F(x) \, dx. \tag{3}$$

The first of these which I owe to Prof. Wilson is implicit in Charlier's work. The second which may fairly be attributed to Aitken may reduce the actual work in many special cases met in practice.

Both of these methods are closely related to the Charlier device of finding polynomials $S_n(x)$ with the bi-orthogonal property

$$\int_{-\infty}^{\infty} S_n(x) f^i(x) \, dx = 0, \qquad i \neq n.$$

The subscript indicates the degree of the polynomial. By means of n of the above relationships, the polynomials can be determined except for a factor of

proportionality. By formal integration of both sides of our expansion we have the Charlier identity

$$a_n = \int_{-\infty}^{\infty} S_n(x)F(x)\,dx/\text{factor of proportionality.}$$

From a theoretical standpoint, this method leaves little to be desired; but in practice the algebraic work increases rapidly with the number of terms to be included in the series.

In the Carver method, the new parent function in question, as well as the function to be approximated, are both expanded in terms of the normal curve, thus almost doubling the numerical calculations. After some differentiation, the members of the Type A family are eliminated yielding in the process the required coefficients in terms of the new parent family. We shall see below how this method may be related to the three above.

3. Useful relationship. First, two simple identities may be presented:

$$\begin{aligned} L_j(f^i) &= (-1)^i L_{j-i}(f), \quad j \geqq i \\ &= 0 \qquad\qquad\quad , \quad j < i. \end{aligned}$$

Given the above assumptions of high contact, this follows immediately from repeated integration by parts.

Remembering that the reduced moments defined just above are the coefficients of the powers of α in the series expansion of the moment generating function

$$M(\alpha; f^i) = \int_{-\infty}^{\infty} e^{\alpha x} f^i(x)\,dx = L_0(f^i) + L_1(f^i)\alpha + L_2(f^i)\alpha^2 + \cdots$$

we have the useful Aitken identity

$$M(\alpha; f^i) = (-1)^i M(\alpha; f)\alpha^i. \tag{4}$$

This, too, is the immediate consequence of repeated integration by parts.

4. Derivation of first method. Formally multiplying each side of (1) by $x^n/n!$ and integrating, we have the formal identity

$$L_n(F) = a_0 L_n(f) - a_1 L_{n-1}(f) + \cdots + (-1)^n a_n L_0(f)\,.$$

This is a "triangular" system of linear equations in the unknown a's. It may be written in matrix terms

$$\begin{bmatrix} L_0(F) \\ L_1(F) \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} = \begin{bmatrix} L_0(f) & 0 & 0 & \cdots \\ L_1(f) & L_0(f) & 0 & \cdots \\ L_2(f) & L_1(f) & L_0(f) & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{bmatrix} \begin{bmatrix} a_0 \\ -a_1 \\ a_2 \\ -a_3 \\ \cdot \\ \cdot \end{bmatrix}.$$

The triangular matrix has the very special property that all of its elements are known as soon as the first column is given. For this reason, as we shall see, it is essentially equivalent to a simple sequence of numbers. This we shall call the *sequence property*. Because of this special form, the above system by simple rearrangement may be written in the modified form

$$\begin{bmatrix} L_0(F) & 0 & \cdots \\ L_1(F) & L_0(F) & \cdots \\ \cdot & \cdot & \\ \cdot & \cdot & \\ \cdot & \cdot & \\ \cdot & \cdot & \end{bmatrix} = \begin{bmatrix} L_0(f) & 0 & \cdots \\ L_1(f) & L_0(f) & \cdots \\ \cdot & \cdot & \\ \cdot & \cdot & \\ \cdot & \cdot & \\ \cdot & \cdot & \end{bmatrix} \begin{bmatrix} a_0 & 0 & 0 \cdots \\ -a_1 & a_0 & 0 \cdots \\ a_2 & -a_1 & a_0 \cdots \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \end{bmatrix}$$

By appropriate definition of symbolism, this may be written in the simple matrix form:

$$L(F) = L(f)\, a(F, f),$$

since multiplication of two triangular, "sequence" matrices is commutative.

It is usually simplest to invert this triangular solution directly as in (2). But if necessary, we may express our answer in the equivalent form

$$a(F, f) = L(F)\, L(f)^{-1}, \tag{5}$$

where the inverse to any special triangular matrix, also possesses the sequence property.

If g is a second parent function with the properties of Section 2, we have the relationship

$$a(F, g) = a(F, f)\, a(f, g)$$

which follows directly from (5). This may be generalized to

$$a(f_1, f_2)\, a(f_2, f_3) \cdots a(f_{n-1}, f_n) = a(f_1, f_n)\,.$$

If F itself is a parent function, we have

$$a(F, f)\, a(f, F) = a(F, F) = I$$

or

$$a(f, F) = a(F, f)^{-1}\,.$$

5. Relation to old methods. In terms of our notation, the Carver method seems to reduce to computing $a(F, f)$ by the relationship

$$a(F, f) = a(F, \phi)\, a(f, \phi)^{-1}$$

where ϕ is the Type A parent function. It involves a doubling of the work of coefficient determination. However, if only a few terms in the expansion are retained, this is of negligible importance.

The Charlier polynomials are clearly summed rows of the matrix product

$$L(f)^{-1} \cdot \begin{bmatrix} 1/1! & 0 & 0 & \cdots \\ 0 & x/1! & 0 & \cdots \\ 0 & 0 & x^2/2! & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{bmatrix}.$$

To know the first n of these polynomials, it is not necessary to derive $n(n+1)/2$ different coefficients. Because of the sequence property, it is only necessary to derive n elements of the first column of $L(f)^{-1}$. These can be expressed in terms of the reduced moments of f, as did Charlier; but the relationships are non-linear and algebraically become tedious for high n. They are better computed from sequence relationships.

The above discussion suggests that the bi-orthogonal relationship between a parent family and suitable polynomials has no deep significance. In particular, there is no essential relationship to least squares as in orthogonal expansions. It does, however, share one important property with orthogonal functions—determination of later coefficients does not affect the earlier ones. But this is a property of all triangular reductions, orthogonal expansions being only special cases of these.

6. Sequence properties. Ordinarily to derive the inverse of an n^2 matrix, n^2 equations must be solved. For our triangular matrices, we need only solve n equations for one column. To each triangular matrix $L(f)$ there corresponds a sequence $\{L_k(f)\}$, which is in fact the first column of the former. Similarly to $L(f)^{-1}$, there corresponds $\{\bar{L}_k(f)\}$; the elements of the latter are defined by the n equations

$$\begin{aligned} &L_0(f)\bar{L}_0(f) = 1 \\ &L_0(f)\bar{L}_1(f) + L_1(f)\bar{L}_0(f) = 0 \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ &\sum_0^n L_k(f)\bar{L}_{n-k}(f) = 0 \end{aligned}$$

But these are precisely the equations involved in the formal inversion of any linear operator system of the form

$$\sum_0^\infty c_k h^k y = z \tag{6}$$

where h is an operator which commutes with a constant, and for which $h^0 = 1$. z is a known function and y unknown. Thus h may be such operators as

$$x,\ d/dx,\ xd/dx,\ E,\ \Delta.$$

A particular solution of (6) is given by the formal expansion

$$y = \sum_0^\infty \tilde{c}_j h^j z$$

where the $\tilde{c}$'s bear the same relationship to the c's as do the $\tilde{L}$'s to the L's.

Such "reciprocal" sequences appear in many branches of applied mathematics. In particular, they arise in the inversion of a power series. If formally,

$$W(\alpha) = \sum_0^\infty S_k \alpha^k$$

then

$$\frac{1}{W(\alpha)} = \sum_0^\infty \tilde{S}_k \alpha^k.$$

Thus, to any triangular matrix with the sequence property, we can formally associate a function $W(\alpha)$ as well as a sequence of numbers. The calculus of multiplication of our triangular matrices clearly "corresponds" to the calculus of multiplication of functions, i.e. if the triangular matrices $T_1, T_2, \cdots T_n$ and $W_1(\alpha), W_2(\alpha), \cdots W_n(\alpha)$ correspond, and $T_n = T_1 \cdot T_2 \cdots T_{n-1}$; then

$$W_n(\alpha) = W_1(\alpha) W_2(\alpha) \cdots W_{n-1}(\alpha).$$

Also, $1/W_i(\alpha)$ corresponds to T_i^{-1}.

7. Moment generating functions. If only for the above reasons and no others, we should be tempted to consider the function formally defined by

$$\sum_0^\infty L_k(f) \alpha^k.$$

But this is precisely the expression for the familiar moment generating function, m. g. f.

$$M(\alpha; f) = \int_{-\infty}^{\infty} e^{\alpha x} f(x)\, dx = \sum_0^\infty L_k(f) \alpha^k.$$

In this way, the method of triangular matrices joins the method used by Aitken for the Type A family. If

$$F(x) \sim \sum_0^\infty a_i f^i(x),$$

and we formally equate moment generating functions of each side, we get

$$M(\alpha; F) = M(\alpha; f) \sum_0^\infty (-1)^i a_i \alpha^i, \tag{7}$$

by means of the Aitken identity (4). Thus $(-1)^i a_i$ equals the coefficient of α^i in the formal expansion of

$$\frac{M(\alpha; F)}{M(\alpha; f)} = M(\alpha; F)M(\alpha; f)^{-1}.$$

Our relationship (2) follows immediately from (7); and by Taylor's expansion in α of $M(\alpha; f)^{-1}$, the identity (3) is quickly realized.

For many problems, the reciprocal of the m. g. f. of $f(x)$ is itself a simple function; to that our triangular equations may be inverted without solving linear equations. Thus where $F(x) = f(x + b)$, we immediately verify Taylor's expansion by use of familiar properties of the m. g. f. under shift of origin.

8. Finite difference expansions. Corresponding to integration by parts, we have the formula

$$\sum_{-\infty}^{\infty} W_i \nabla^k V_i = (-1) \sum_{-\infty}^{\infty} \Delta W_i \nabla^{k-1} V_i = (-1)^2 \sum_{-\infty}^{\infty} \Delta^2 W_i \nabla^{k-2} V_i, \quad \text{etc.},$$

provided "high contact" properties are assumed. ∇ and Δ are receding and advancing differences respectively. Recalling the familiar property of "reduced factorial" polynomials, ${}^k x$, we have

$$\sum_{-\infty}^{\infty} {}^j x \nabla^k f(x) = (-1)^k \sum_{-\infty}^{\infty} {}^{j-k} x f(x) \qquad j \geqq k$$

$$= 0 \qquad j < k,$$

or

$$Q_j(\nabla^k f) = (-1)^k Q_{j-k}(f) \qquad j \geqq k$$

$$= 0 \qquad j < k,$$

where

$$Q_j(g) = \sum_{-\infty}^{\infty} \frac{x(x-1)(x-2)\cdots(x-j+1)}{j!} g(x).$$

In the expansion

$$F(x) \sim a_0 f(x) + a_1 \nabla f(x) + a_2 \nabla^2 f(x) + \cdots,$$

the a's obey laws identical to (2) and (3) where reduced factorial moments are substituted for the reduced L moments, and the f. m. g. f.

$$\sum_{-\infty}^{\infty} f(x)(1 + \alpha)^x,$$

for the ordinary m. g. f.

9. Convergence. All of the above relationships are purely formal, without regard to convergence. The last is a difficult subject, and little discussed in the statistical literature, since applications of *G-C* series have been almost entirely concerned with empirical frequency curve fitting in which mathematical con-

vergence does not enter. Actually in the scanty treatments of the subject there has arisen a confusion between the Type A *G-C* expansion, which equates moments, and the expansion of a function in orthogonal Hermite functions. These are not unrelated, but nevertheless they are distinct. This is well recognized in the purely mathematical literature, but hardly at all in the literature of statistics and physics.

The series differ by an irremovable factor of 2. If the Type A functions are written as

$$H_i(x)e^{-x^2},$$

then the Hermite functions will take the form

$$H_i(x)e^{-\frac{1}{2}x^2}.$$

where the H's are Hermite polynomials suitably normalized. Unfortunately the *G-C* series often diverges when the H series converges. Thus, the statistically interesting Cauchy distribution can be expanded in an H series; but since it possesses no finite higher moments, the *G-C* series cannot even be defined.

It is not hard to show that the *G-C* expansion of F in terms of a Type A function $f(x)$, is equivalent to an H expansion of $Ff^{-\frac{1}{2}}$ in terms of the H family $f^{\frac{1}{2}}$. It is sufficient for convergence in the mean of the last expansion that $Ff^{-\frac{1}{2}}$ be of integrable square or belong to L^2. This means that the *G-C* type A expansion will be valid if $Ff^{-\frac{1}{2}}$ is well behaved, not simply if F is well behaved. For F a histogram as is often the case in practise, no difficulties of convergence arise, although rapid convergence may be another matter. Nevertheless, many well behaved F's will not pass the more strict test. The reader is referred to the last five titles in the bibliography for mathematical discussions of this problem.

The above discussion holds only for the Type A expansion. There remains the very difficult problem of convergence conditions in the more general case. No immediate generalization suggests itself, except the application of the results of the "moment problem." However, this must be handled with delicacy, since the partial sums of the series may actually become negative over some range.

BIBLIOGRAPHY

References

[1] J. P. Gram, "Ueber die Entwicklung reeler Functionen in Reihen mittelst der Methode der Kleinsten Quadrate," *Journal für die Reine und Angewandte Mathematik*, Vol. 94 (1882), pp. 41–73.

[2] H. Bruns, *Wahrscheinlichkeitsrechnung und Kollektivmasslehre*, B. G. Teubner, Leipzig, 1906.

[3] C. V. L. Charlier, "Über die Darstellung willkürlicher Funktionen," *Arkiv för Matematik, Astronomi och Fysik (utgivet af k. svenska vetenskapsakademien)*, Vol. 2 (1905), 35 pp.

[4] H. C. Carver, Chapter VII, *Handbook of Mathematical Statistics*, edited by H. L. Rietz, Cambridge, Massachusetts, 1924.

[5] Emeterio Roa, "A number of new generating functions with application to statistics," 1923.

[6] A. C. AITKEN, *Statistical Mathematics*, London, 1939.
[7] V. ROMANOVSKY, "Generalisation of some Types of the Frequency Curves of Professor Pearson," *Biometrika*, Vol. 16 (1924), pp. 106–116.
[8] DUNHAM JACKSON, *Fourier Series and Orthogonal Polynomials*, (Carus Mathematical Monograph No. 6), The Mathematical Association of America, Oberlin, Ohio, 1941.
[9] E. H. HILDEBRANDT, "Systems of polynomials connected with the Charlier expansions and the Pearson differential and difference equations," *Annals of Math. Stat.*, Vol. II, 1931, pp. 379–439.

Further Literature

1. W. MYLLER-LEBEDEFF, "Die Theorie der Integralgleichungen in Anwendung auf einige Reihenentwicklungen," *Math. Annalen*, Vol. 64 (1907), pp. 388–416.
2. E. HILLE, "A class of reciprocal functions," *Annals of Math.*, Vol. 27 (1926), pp. 427–464. (Contains selected bibliography.)
3. M. H. STONE, "Developments in Hermite polynomials," *Annals of Math.*, Vol. 29 (1927), pp. 1–13.
4. W. E. MILNE, "On the degree of convergence of the Gram-Charlier series," *Trans. Amer. Math. Soc.*, Vol. 31 (1929), pp. 422–444.
5. *Bibliography on Orthogonal Polynomials*, National Research Council of the National Academy of Sciences, Washington, D. C., 1940.

A METHOD OF TESTING THE HYPOTHESIS THAT TWO SAMPLES ARE FROM THE SAME POPULATION

By Harold C. Mathisen

Princeton University

1. Introduction. There are many cases in testing whether two samples are from the same population in which no assumption about the distribution function of the population can be made except that it is continuous. A. Wald and J. Wolfowitz, [1], have developed a method of testing the hypothesis that two samples come from the same population based on certain kinds of runs of the elements from each sample in the combined ordered sample. W. J. Dixon, [2], has introduced a criterion for testing the same hypothesis based on the number of elements of the second sample falling between each successive pair of ordered values in the first sample.

The problem considered here is that of devising a simple method of testing the hypothesis that two samples come from the same population, based on medians and quartiles, given only that the distribution function of the population is continuous. The simplest method may be described briefly as follows. We observe the number of elements, m_1, in the second sample whose values are lower than the median of the first sample. Since the distribution of m_1 is independent of the population distribution, we are able to compute significance points from the distribution of m_1. These points may then be used for testing the hypothesis at a given significance level. This will be referred to as the case of two intervals.

This method may be easily extended to the case of any number of intervals. In this note we shall consider the extension to four intervals by using the median and the two quartiles of the first sample to establish four intervals into which the elements of the second sample may fall. Then, if the second sample is of size $4m$, it will be shown that, under the hypothesis that the two samples come from the same population, $\frac{1}{4}$ of the second sample, or m elements will be expected to fall in each interval. Let the number in the second sample which actually fall in each interval be m_1, m_2, m_3, and m_4 respectively. The test function here proposed is,

$$C = \frac{(m_1 - m)^2 + (m_2 - m)^2 + (m_3 - m)^2 + (m_4 - m)^2}{9m^2}, \tag{1}$$

where $9m^2$ is a constant, which forces C to lie on the interval 0 to 1. If the m_i, $(i = 1, 2, 3, 4)$, have values quite different from their expected value m, it is apparent that C will be large. Therefore the greater the value of C the more doubtful is the hypothesis that the two samples come from the same population. Significance values of C will be computed for several sample sizes. The question of whether C is the "best" four-interval criterion for testing the hypothesis that two samples come from the same continuous distribution is an open one

which would depend for its answer on an extensive power function analysis. We shall not go into this analysis, however, but shall use C on intuitive grounds. This case will be referred to as the case of four intervals. The extension of the method of the case of four intervals to any number of intervals presents no new difficulties in derivation, however we shall confine our attention to the cases of two and four intervals.

2. The case of two intervals. Suppose $f(x)$ is a continuous distribution function with probability element $f(x)\,dx$. Let us draw a sample of size $2n + 1$ from a population having this probability element. Let the elements in the sample be $x_1, x_2, \cdots, x_{2n+1}$ ordered from least to greatest. The median of this sample will be x_{n+1}. Now consider a second sample of size $2m$, and let m_1 be the number of observations, whose values are less than x_{n+1}. We call $m_2 = 2m - m_1$ the number of elements in the second sample greater than x_{n+1}.

Let $p = \int_{-\infty}^{x_{n+1}} f(x)\,dx$ be the probability of an observation having a value less than x_{n+1}. Then the probability of an element having a value greater than x_{n+1} is $(1 - p)$. Thus we have the relation $f(x_{n+1})\,dx_{n+1} = dp$. The probability law of the median, x_{n+1} given by the multinomial law[1] is

$$P_r(x_{n+1}) = \frac{(2n+1)!}{n!1!n!}\,p^n(1-p)^n\,dp. \tag{2}$$

The conditional probability law of m_1, given x_{n+1}, is then

$$P_r(m_1 \mid x_{n+1}) = \frac{(2m)!}{m_1!(2m-m_1)!}\,p^{m_1}(1-p)^{2m-m_1}. \tag{3}$$

From this it follows that the joint probability law of x_{n+1} and m_1 is the product of (2) and (3) or

$$P_r(m_1, x_{n+1}) = \frac{(2n+1)!(2m)!}{n!n!m_1!(2m-m_1)!}\,p^{n+m_1}(1-p)^{n+2m-m_1}\,dp. \tag{4}$$

We may integrate (4) with respect to p from 0 to 1 as a Beta Function, leaving the distribution function of m_1 independent of the population probability element $f(x)\,dx$. We get for the distribution of m_1,

---

[1] The multinomial law may be stated briefly as follows:

If a trial results in one and only one of the mutually exclusive events $E_1, E_2, \cdots, E_k$, the probability P that in a total of n trials, n_1 will result in E_1, n_2 in E_2, $\cdots$, n_k in E_k, $\left(\sum_1^k n_i = n\right)$, is given by

$$P = \frac{n!}{n_1!\,n_2!\cdots n_k!}\,p_1^{n_1}p_2^{n_2}\cdots p_k^{n_k}$$

where $p_1, p_2, \cdots, p_k$, $\left(\sum_1^k p_i = 1\right)$, are the probabilities of a single trial resulting in $E_1, E_2, \cdots, E_k$ respectively.

$$(5) \qquad P_r(m_1) = \frac{(2n+1)!(2m)!(n+m_1)!(n+2m-m_1)!}{n!n!m_1!(2m-m_1)!(2n+1+2m)!}.$$

From (5) a simple recursion relation between $P_r(m_1)$ and $P_r(m_1+1)$ may be determined from which the probabilities of various values of m may be rapidly computed. For large samples it can be shown that under certain regularity conditions, the ratio, $[m_1 - E(m_1)]/\sigma_{m_1}$ may be approximated by the normal distribution[2] with zero mean and unit variance. The derivation is similar to that of the four-interval case, which is taken up in greater detail. It will be found by the use of (4) that the expected value of m_1 is m, and the variance of m_1 is $m + \frac{m(2m-1)(n+2)}{2n+3} - m^2$. Using this information, values of m_1 for various

TABLE I

The Case of Two Intervals

Lower and upper .01 and .05 percentage points for the distribution of m_1

Sample sizes		Critical values of m_1			
		Lower		Upper	
First $2n+1$	Second $2m$	$m_{1(.01)}$	$m_{1(.05)}$	$m_{1(.05)}$	$m_{1(.01)}$
11	10		1	9	
41	40	10	12	28	30
101	100	34	38	62	66
101	200	72	80	120	128
201	200	77	84	116	123
201	400	160	181	219	240
401	400	167	177	223	233
401	800	353	367	433	447
1001	1000	448	463	537	552

significance levels may be computed. The .01 and .05 percentage points of m_1 for several sample sizes are given in Table I. The values for sample sizes of 10 and 40 are computed directly from the probability law, while the larger samples have limits computed by the normal approximation. Thus for two samples of size 101 and 100, respectively, a value of m_1 less than 38 would be significant at the .05 level. Similarly, at the upper .05 level, the hypothesis would be rejected if a value of m_1 were obtained which was greater than 62. The necessity for the upper limits could easily be eliminated by testing with respect to the smaller of m_1 and m_2. However, for completeness, the upper percentage points

[2] This statement may be proved by showing that as $m, n \to \infty$ such that $m/n = constant$, the limit of the moment generating function for the ratio is identical with the moment generating function of the normal distribution with zero mean and unit variance.

are included to show the range of values of m_1 in which the hypothesis that the two samples come from the same population may be accepted.

3. The case of four intervals. If we let the first sample of size $4n + 3$ be designated by $(x_1, x_2, \cdots x_{4n+3})$, assumed drawn from a population with probability element $f(x)\,dx$ and ordered from least to greatest, then the range of x may be divided into four intervals by x_{n+1}, x_{2n+2}, and x_{3n+3}. The probability element of x_{n+1}, x_{2n+2}, x_{3n+3} is

$$\frac{(4n+3)!}{n!n!n!n!}\left(\int_{-\infty}^{x_{n+1}} f(x)\,dx\right)^n\left(\int_{x_{n+1}}^{x_{2n+2}} f(x)\,dx\right)^n\left(\int_{x_{2n+2}}^{x_{3n+3}} f(x)\,dx\right)^n\left(\int_{x_{3n+3}}^{\infty} f(x)\,dx\right)^n$$
$$\cdot f(x_{n+1})\,dx_{n+1} f(x_{2n+2})\,dx_{2n+2} f(x_{3n+3})\,dx_{3n+3}.$$

TABLE II

The Case of Four Intervals

.95 and .99 percentage points for the distribution of C

Sample sizes				$C_{.95}$	$C_{.99}$
First	Second				
$4n+3$	$4m$	n	m		
15	12	3	3	.446	.582
63	60	15	15	.113	.161
103	100	25	25	.072	.102

Let

$$\int_{-\infty}^{x_{n+1}} f(x)\,dx = p_1, \quad \int_{x_{n+1}}^{x_{2n+2}} f(x)\,dx = p_2, \quad \int_{x_{2n+2}}^{x_{3n+3}} f(x)\,dx = p_3, \quad \int_{x_{3n+3}}^{\infty} f(x)\,dx = p_4.$$

The probability element of p_1, p_2, p_3, and p_4 is

$$p_r(x_{i(n+1)}) = \frac{(4n+3)!}{n!1!n!1!n!1!n!}\,p_1^n\,p_2^n\,p_3^n\,p_4^n\,dp_1\,dp_2\,dp_3. \tag{6}$$

Now let us consider the second sample, $(x_1', x_2', \cdots x_{4m}')$, of size $4m$. Let the number of observations falling in each of the preassigned intervals be m_i, $(i = 1, 2, 3, 4)$, where $m_4 = 4m - m_1 - m_2 - m_3$. The conditional probability of the m_i, given the values of $x_{i(n+1)}$ is also determined by the multinomial law.

$$P_r(m_i \mid x_{i(n+1)}) = \frac{(4m)!}{m_1!m_2!m_3!m_4!}\,p_1^{m_1}p_2^{m_2}p_3^{m_3}p_4^{m_4}. \tag{7}$$

The joint distribution of the p_i and the m_i is then

$$P_r(x_{i(n+1)}, m_i) = \frac{(4n+3)!\,(4m)!}{(n!)^4 m_1!m_2!m_3!m_4!}\,p_1^{n+m_1}p_2^{n+m_2}p_3^{n+m_3}p_4^{n+m_4}\,dp_1\,dp_2\,dp_3. \tag{8}$$

To obtain the distribution of the m_i alone, the p_i will be integrated out by the Dirichlet Integral[3] formula, giving a distribution which is clearly independent of the population distribution function $f(x)$.

$$(9) \qquad P_r(m_i) = \frac{(4n+3)!\,(4m)!\,(n+m_1)!\,(n+m_2)!\,(n+m_3)!\,(n+m_4)!}{(n!)^4 m_1!\,m_2!\,m_3!\,m_4!\,(4m+4n+3)!}.$$

To find the expected value of the m_i, the probability law of m_1 will first be derived. The probability function for the value of x_{n+1} is

$$(10) \qquad P_r(x_{n+1}) = \frac{(4n+3)!}{1!\,n!\,(3n+2)!}\, p_1^n (1-p_1)^{3n+2}\, dp_1 .$$

Then we have the conditional probability

$$(11) \qquad P_r(m_1 \mid x_{n+1}) = \frac{(4m)!}{m_1!(4m-m_1)!}\, p_1^{m_1}(1-p_1)^{4m-m_1},$$

and

$$(12) \qquad P_r(x_{n+1}, m_1) = \frac{(4n+3)!\,(4m)!}{n!\,(3n+2)!\,m_1!\,(4m-m_1)!}\, p_1^{n+m_1}(1-p_1)^{3n+2+4m-m_1}\, dp_1 .$$

To obtain the expected value of m_1, the joint distribution of m_1 and p_1 is multiplied by m_1, summed on m_1 from 0 to $4m$, and integrated on p_1 from 0 to 1.

$$(13) \qquad E(m_1) = \frac{(4n+3)!}{n!\,(3n+2)!} \int_0^1 p_1^n (1-p_1)^{3n+2} \cdot \left[\sum_0^{4m} m_1 \frac{(4m)!}{m_1!\,(4m-m_1)!}\, p_1^{m_1}(1-p_1)^{4m-m_1} \right] dp_1 .$$

This interchange of the order of integration and summation is clearly valid. The quantity in brackets will be recognized as the first moment of the binomial distribution, $(p_1+q)^{4m}$ where $q = 1 - p_1$. Therefore we have

$$(14) \qquad E(m_1) = \int_0^1 4mp_1 f(p_1)\, dp_1 = 4mE(p_1).$$

$E(p_1)$ and the higher moments of p_1 are found in the usual way by integrating the distributions as Beta Functions. From this we see that the expected value of m_1 is m. By repeating these operations on m_2, m_3, and m_4, it can be seen that $E(m_i) = m$, which also validates the statement made in the introduction.

---

[3] A discussion of the Dirichlet Integral may be found in Woods—*Advanced Calculus*, p. 167. It may be stated as follows for the problem in which we are interested

$$\iiint x^{l-1} y^{m-1} z^{n-1} (1-x-y-z)^{r-1}\, dx\, dy\, dz = \frac{\Gamma(l)\Gamma(m)\Gamma(n)\Gamma(r)}{\Gamma(l+m+n+r)},$$

where we integrate over the region bounded by $x + y + z = 1$, and the three coordinate planes.

We have previously presented the criterion (1).

The next problem is to find a distribution function to which the distribution of C may be fitted. A reasonable choice appears to be the Pearson Type I curve.

$$f(x) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} x^{r-1}(1-x)^{s-1}. \tag{15}$$

The distribution of C is fitted by equating the first two moments of the two distributions and solving for the constants r and s of the Type I distribution. Using the theorem that the mean value of the sum of variates is equal to the sum of their mean values, we have

$$E(C) = \frac{1}{9m^2}[E(m_1^2) + E(m_2^2) + E(m_3^2) + E(m_4^2) - 4m^2]. \tag{16}$$

Also the second moment may be written as

$$\begin{aligned} E(C^2) = \frac{1}{81m^4}[&E(m_1^4) + E(m_2^4) + E(m_3^4) + E(m_4^4) + 16m^4 + 2E(m_1^2 m_2^2) \\ &+ 2E(m_1^2 m_3^2) + 2E(m_1^2 m_4^2) + 2E(m_2^2 m_3^2) + 2E(m_2^2 m_4^2) \\ &+ 2E(m_3^2 m_4^2) - 8m^2\{E(m_1^2) + E(m_2^2) + E(m_3^2) + E(m_4^2)\}]. \end{aligned} \tag{17}$$

The expected value of m_1^2 is found in the same manner as $E(m_1)$ and here also it can be shown that the $E(m_i^2)$ are all equal. The same procedure holds for $E(m_i^4)$.

$$E(m_i^2) = m + \frac{m(4m-1)(n+2)}{4n+5},$$

$$\begin{aligned} E(m_i^4) = m &+ \frac{7m(4m-1)(n+2)}{4n+5} + \frac{6m(4m-1)(4m-2)(n+3)(n+2)}{(4n+6)(4n+5)} \\ &+ \frac{m(4m-1)(4m-2)(4m-3)(n+4)(n+3)(n+2)}{(4n+7)(4n+6)(4n+5)}. \end{aligned} \tag{18}$$

By using the moment generating function of the trinomial distribution, the $E(m_i^2 m_j^2)$ may also be found in a similar manner.

$$\begin{aligned} E(m_i^2 m_j^2) = &\frac{m(4m-1)(n+1)}{4n+5} + \frac{2m(4m-1)(4m-2)(n+1)(n+2)}{(4n+6)(4n+5)} \\ &+ \frac{m(4m-1)(4m-2)(4m-3)(n+2)(n+1)(n+2)}{(4n+7)(4n+6)(4n+5)}. \end{aligned} \tag{19}$$

As a result we have

$$E(C) = \frac{4}{9m} + \frac{4(4m-1)(n+2)}{9m(4n+5)}. \tag{20}$$

Let $E(C) = A$ to simplify later relations to be computed. Finally

$$E(C^2) = \frac{4}{81m^3}\Bigg[1 + \frac{7(4m-1)(n+2)}{4n+5} + \frac{6(4m-1)(4m-2)(n+3)(n+2)}{(4n+6)(4n+5)} + \frac{(4m-1)(4m-2)(4m-3)(n+4)(n+3)(n+2)}{(4n+7)(4n+6)(4n+5)} + 4m^3 + \frac{3(4m-1)(n+1)}{4n+5} + \frac{6(4m-1)(4m-2)(n+1)(n+2)}{(4n+6)(4n+5)} + \frac{3(4m-1)(4m-2)(4m-3)(n+2)^2(n+1)}{(4n+7)(4n+6)(4n+5)} - 8m^2 - \frac{8m^2(4m-1)(n+2)}{4n+5}\Bigg]. \tag{21}$$

To simplify later relations we let $E(C^2) = B$.

The first two moments of the Type I distribution are easily found to be

$$\mu_1 = \frac{r}{r+s} = A \qquad \mu_2 = \frac{\mu_1(r+1)}{(r+s+1)} = B. \tag{22}$$

Solving these two simultaneous equations for r and s,

$$r = \frac{B-A}{A-\frac{B}{A}} \qquad s = \frac{r}{A} - r. \tag{23}$$

A number of percentage points for the Type I distribution have been computed by Miss Catherine Thompson, [3]. Using these limits, the hypothesis may be accepted or rejected as to whether or not the two samples come from the same population.

Table II shows the .95 and .99 percentage points of C for three sample sizes.

4. Summary. The problem considered here is that of devising a simple method of testing the hypothesis that two samples are from identical populations having continuous distribution functions. It may be summarized briefly as follows. The first sample is used to establish any desired number of intervals into which the observations of the second sample may fall. A test criterion is proposed which is based on the deviations of the numbers of elements of the second sample which fall in the intervals from the expected values of the respective numbers. Two cases are discussed, that of two intervals and that of four intervals, making use of the median and quartiles in the first sample to determine the intervals. Tables of 1% and 5% points for several sample sizes of both cases are given.

REFERENCES

[1] A. Wald and J. Wolfowitz, *Annals of Math. Stat.*, Vol. 11 (1940), p. 147.
[2] W. J. Dixon, *Annals of Math. Stat.*, Vol. 11 (1940), p. 199.
[3] Catherine Thompson, *Biometrika*, Vol. 32 (1941), p. 151.

NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

NOTE ON THE INDEPENDENCE OF CERTAIN QUADRATIC FORMS

By Allen T. Craig

University of Iowa

Various approaches to the problem of the independence of quadratic forms in normally and independently distributed variables have been made by R. A. Fisher, Cochran, Madow and others. It is the purpose of this note to point out a few simple propositions which, in so far as the writer is aware, have not had specific mention in the literature.

1. Independence of certain quadratic forms. Theorem 1: *A necessary and sufficient condition that two real symmetric quadratic forms, in n normally and independently distributed variables, be independent in the probability sense is that the product of the matrices of the forms be zero.*

Let the chance variable x be normally distributed with mean zero and unit variance. Let $x_1, x_2, \cdots, x_n$ be n independent values of x and let A and B be two real symmetric matrices, each of order n. Write $Q_1 = \Sigma\Sigma a_{ij}x_ix_j$ and $Q_2 = \Sigma\Sigma b_{ij}x_ix_j$ where $\| a_{ij} \| = A$ and $\| b_{ij} \| = B$. It is well known that the generating function of the moments of the joint distribution of Q_1 and Q_2 can be written

$$G(\lambda, \lambda') = | I - \lambda A - \lambda' B |^{-\frac{1}{2}},$$

so that

$$| I - \lambda A - \lambda' B | = | I - \lambda A | \, | I - \lambda' B |, \tag{1}$$

for all real values of λ and λ', is necessary and sufficient for the independence of Q_1 and Q_2.

If Q_1 and Q_2 are independent, then (1), being true for all real values of λ and λ', is in particular true for $\lambda = \lambda'$. Thus

$$| I - \lambda(A + B) | \equiv | I - \lambda A | \, | I - \lambda B |. \tag{2}$$

Denote by r_1, r_2 and $r \leq r_1 + r_2$ respectively the ranks of A, B and $A + B$. Then $r = r_1 + r_2$ since (2) expresses the identity of two polynomials in λ of degrees r and $r_1 + r_2$.

Further, if we write

$$| I - \lambda A | = (1 - \lambda p_1) \cdots (1 - \lambda p_{r_1}),$$

$$| I - \lambda B | = (1 - \lambda q_1) \cdots (1 - \lambda q_{r_2}),$$

and $| I - \lambda(A + B) | = (1 - \lambda s_1) \cdots (1 - \lambda s_{r_1+r_2})$, then, because the factorization of polynomials is unique, each s_j can be paired with one of the numbers $p_1, \cdots, p_{r_1}, q_1, \cdots, q_{r_2}$. Thus, if Q_1 and Q_2 are independent, the rank of $A + B$ is the sum of the ranks of A and B, and the non-zero roots of the characteristic equation of $A + B$ are those of the characteristic equation of A together with those of the characteristic equation of B. There exists an appropriately chosen orthogonal matrix L of order n such that $L'(A + B)L$, L' being the conjugate of L, is a matrix with the reciprocals of the numbers $p_1, \cdots, p_{r_1}, q_1, \cdots, q_{r_2}$ on the principal diagonal and zeros elsewhere. Then $L'AL$ and $L'BL$ have no overlapping non-zero elements and $L'ALL'BL = 0$. But $L' = L^{-1}$, the inverse of L. Hence, upon multiplying both members of the preceding equation on the right by L' and on the left by L, we have $AB = 0$. Since $A = A'$ and $B = B'$, likewise $BA = 0$.

Conversely, suppose $AB = 0$. Then the matrix $(I - \lambda A)(I - \lambda' B) = I - \lambda A - \lambda' B$. These matrices being equal, their determinants are equal and the condition (1) for the independence of Q_1 and Q_2 is satisfied.

The theorem is readily extended to the case of the mutual independence of any finite number of such quadratic forms.

The product of a non-singular matrix and a matrix of rank R is a matrix of rank R. Hence, every non-singular quadratic form of the kind here discussed is correlated with every non-identically vanishing quadratic form in the same variables.

2. Conditions for independent Chi-Square distributions. The preceding theorem enables one to determine, by multiplication of matrices, whether real symmetric quadratic forms in normally and independently distributed variables are themselves independent in the probability sense. The following theorem affords a simple test as to whether the distributions are of the Chi-Square type.

THEOREM 2: *Necessary and sufficient conditions that each of two real symmetric quadratic forms, in n normally and independently distributed variables with mean zero and unit variance, be independently distributed as is Chi-Square, are that the product of the matrices of the forms be zero and that each matrix equal its own square.*

If Q_1 and Q_2 are independently distributed as is Chi-Square, then $AB = 0$ and each of the non-zero roots of the characteristic equations of A and B is $+1$. For an appropriately chosen orthogonal matrix L, of order n, $L'AL$ is a matrix with r_1 elements on the principal diagonal $+1$, all other elements being zero. For such a matrix it is seen that $(L'AL)(L'AL) = L'A^2L = L'AL$ and $A^2 = A$. A similar argument shows that $B^2 = B$.

Conversely, if $AB = 0$, then Q_1 and Q_2 are independent. Further, if $A^2 = A$ and $B^2 = B$, each of the non-zero roots of the characteristic equations of A and B is $+1$. This follows from the fact that the roots of the characteristic equation of the square of any matrix are themselves the squares of the roots of the

characteristic equation of that matrix. Since A and B are real and symmetric, the roots under consideration are real. Thus Q_1 and Q_2 have independent Chi-Square distributions with r_1 and r_2 degrees of freedom respectively.

This theorem can likewise be extended to any finite number of these quadratic forms.

Of special interest is the case of, say k, quadratic forms for which the sum of the k matrices is the identity matrix. Thus $A_1 + A_2 + \cdots + A_k = I$. By Theorem 1, it is both necessary and sufficient for the mutual independence of the k forms that $A_u A_v = 0$, $u \neq v$.

Now

$$A_i = I - A_1 - \cdots - A_{i-1} - A_{i+1} - \cdots - A_j - \cdots - A_k$$

and

$$A_i A_j = A_j - A_1 A_j - \cdots - A_{i-1} A_j - A_{i+1} A_j - \cdots - A_j^2 - \cdots - A_k A_j ,$$

so that $A_j = A_j^2$. In this particular case it is to be seen that the mutual independence of the forms implies that their several distributions are of the Chi-Square type.

A CHARACTERIZATION OF THE NORMAL DISTRIBUTION

By Irving Kaplansky

Harvard University

In 1925 R. A. Fisher gave a geometric derivation of the joint distribution of mean and variance in samples from a normal population (*Metron*, Vol. 5, pp. 90–104). On examining the argument however, we find that an (apparently) more general result is actually established: if $f(x_1) \cdots f(x_n)$ is a function $g(m, s)$ of the sample mean m and standard deviation s, then the probability density of m and s in samples of n from the population $f(x)$ is $g(m, s)s^{n-2}$. This condition on $f(x)$ is of course satisfied if $f(x)$ is normal; in this note we shall conversely show that for $n \geqq 3$ it characterizes the normal distribution. In the proof it will be assumed that $g(m, s)$ possesses partial derivatives of the first order, although a weaker assumption would probably suffice.

Let us for the moment restrict the variables x_i to values such that $f(x_i) > 0$. After a change of notation we have

$$\phi(x_1) + \cdots + \phi(x_n) = h(u, v),$$

where $\phi = \log f$, $u = x_1 + \cdots + x_n$, $v = \frac{1}{2}(x_1^2 + \cdots + x_n^2)$. A differentiation yields

$$\phi'(x_i) = h_u + h_v x_i .$$

Solving two of these equations for h_v, we find

$$h_v = \frac{\phi'(x_i) - \phi'(x_j)}{x_i - x_j}, \qquad (i \neq j), \tag{1}$$

and, for $n \geqq 3$, it follows that the right member of (1) is a constant, say 2A. Then

$$\phi'(x_i) - 2Ax_i = \phi'(x_j) - 2Ax_j = a \text{ constant } B.$$

$$\phi(x) = Ax^2 + Bx + C.$$

We now have $f(x) = e^{\phi(x)}$ whenever $f(x) > 0$; but since $f(x)$ is continuous, this implies $f(x) = e^{\phi(x)}$ everywhere.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of general interest

Personal Items

Dr. Holbrook Working has been appointed Chief Statistical Consultant on Industrial Processes and Products in the Office of Production Research and Development of the War Production Board.

Professor Harold Hotelling of Columbia University was the official representative of the Institute of Mathematical Statistics at the Copernican Quadricentennial Celebration which was held in New York City on May 24.

Dr. Edward B. Olds has taken a position with the Curtiss-Wright Corporation.

Dr. Nilan Norris is a Sergeant with the Fourth Statistical Control Unit of the Fourth Air Force with headquarters at San Francisco, California.

Dr. Edward Helly is with the Signal Corps Training Program at Illinois Institute of Technology.

Dr. C. W. Cotterman is in the United States Army at Camp Grant, Illinois.

Mr. M. D. Bingham has been commissioned an Ensign in the United States Naval Reserve and is stationed at Fort Schuyler, New York.

Lt. George W. Petrie, USNR, is teaching in the Midshipmen's School at Notre Dame, Ind.

New Members

The following persons have been elected to membership in the Institute:

Arias B., Jorge Civ. Eng. (Guatemala) Eng., Rural Electrification Administration, 420 Locust St., St. Louis, Mo.

Bailey, A. L. B.S. (Michigan) Stat., American Mutual Alliance, 60 East 42 St., New York, N. Y.

Becker, Harold W. Instr., Mare Island Trainee School. *126 Benson Ave., Vallejo, Calif.*

Bernstein, Shirley R. B.S. (Carnegie Inst. Tech.) Res. Asst., United Steelworkers of America, Pittsburgh, Pa. *5501 Beverly Pl.*

Bickerstaff, Asst. Prof. Thomas A. M.A. (Mississippi) Univ. of Miss., University, Miss.

Birnbaum, Asst. Prof. Z. William. Ph.D. (Lwow) Univ. of Wash., Seattle, Wash.

Brumbaugh, Prof. Martin A. Ph.D. (Pennsylvania) Univ. of Buffalo, Buffalo, N. Y.

Burrows, Glenn L. B.A. (Michigan State Coll.) Instr., Michigan State Coll., East Lansing, Mich.

Cohen, Jozef B. B.S. (Chicago) Sage Fellow in Psychology, Cornell Univ., Ithaca, N. Y.

Cope, Asso. Prof. T. Freeman. Ph.D. (Chicago) Queens College, Flushing, N. Y.

Cudmore, Sedley A. M.A. (Oxford) Stat., Dominion Bur. of Stat., Ottawa, Canada.

Cureton, Edward E. Ph.D. (Columbia) Sr. Personnel Technician, War Dept., *RFD 1, Tauxemont, Alexandria, Va.*

De Castro, Prof. Lauro S. V. Civ. Eng. (Escola Nacional de Enginharia) Catholic Univ., Rio de Janeiro, Brazil. *62 rua David Campista.*

Edwards, G. D. A.B. (Harvard) Dir. of Quality Assurance, Bell Telephone Laboratories, 463 West St., New York, N. Y.

Gifford, Kenneth R. Student, Mass. Inst. Tech., Cambridge, Mass. *97 Bay State Rd., Boston, Mass.*

Gottfried, Bert A. A.M. (Columbia) Stat. Clerk, 4300 Kaywood Dr., Mt. Ranier, Md.

Hamilton, Prof. Thomas R. Ph.D. (Columbia) Texas A. & M. Coll., College Station, Tex.

Heide, J. D. M.S. (Iowa) Stat., U. S. Rubber Co., *1324 Altoona Ave., Eau Claire, Wisc.*

Hilfer, Irma. M.A. (Columbia) Actuary, N. Y. C. Board of Transportation, *165 W. 97 St., New York, N. Y.*

Howell, John M. B.A. (UCLA) Stat., Northrop Aircraft Inc., Hawthorne, Calif. *4140 W. 63 St., Los Angeles, Calif.*

Hurwicz, Leonid. L.L.M. (Warsaw) Res. Asso., Cowles Comm., Univ. of Chicago, Chicago, Ill.

Kendall, Maurice G. M.A. (Cambridge) Stat., Chamber of Shipping of the United Kingdom, *Richmond House, Aldenham Rd., Bushey, Eng.*

Klein, Lawrence R. B.A. (California) Teaching Fellow, Mass. Inst. Tech., Cambridge, Mass.

Kuznets, George M. Ph.D. (California) Instr., Giannini Foundation, Univ. of Calif., Berkeley, Calif.

Landau, H. G. M.S. (Carnegie Inst. Tech.) Stat. Analyst, War Dept., Washington, D. C. *2408 20 St., N.E.*

Langmuir, Charles R. Ed.M. (Harvard) Carnegie Foundation. *437 West 59 St., New York, N. Y.*

Levy, Henry C. L.L.B. (Fordham) Instr., N. Y. C. C., New York, N. Y. *600 West 116 St.*

Li, Jerome C. R. B.S. (Nanking) Student, Iowa State Coll., Ames, Iowa. *2184 Lincoln Way.*

Lieberman, Jacob E. B.S. (Brooklyn Coll.) Jr. Stat., Census Bureau, Washington, D. C. *2422 14 St., N. E.*

Martin, Margaret P. M.A. (Minnesota) Instr., Columbia Univ., New York, N. Y. *1230 Amsterdam Ave.*

Nash, Stanley W. B.A. (Coll. of Puget Sound) San Joaquin Experimental Range, O'Neals, Calif.

Norton, Horace W. Ph.D. (London) Sr. Meteorologist, U. S. Weather Bur., Washington, D. C. *3118 North First Rd., Arlington, Va.*

Olds, Edward B. Ph.D. (Pittsburgh) Stat., Curtiss-Wright Corp. *298 Niagra Falls Blvd., Buffalo, N. Y.*

Preston, Bernard. C.P.A., *103 Park Ave., New York, N. Y.*

Rosenblatt, David. B.S. (Coll. City of N. Y.) Asst. Stat., *1422 Whittier St., N. W., Washington, D. C.*

Sard, Asst. Prof. Arthur. Ph.D. (Harvard) Queens College, Flushing, N. Y. *146-19 Beech Ave.*

Schapiro, Anne. B.A. (Bryn Mawr) Jr. Analyst, Institute of Applied Econometrics, *350 W. 57 St., New York, N. Y.*

Simpson, William B. Grad. Student, Columbia Univ., New York, N. Y.

Springer, Melvin D. M.S. (Illinois) Asst. Instr., Univ. of Illinois, Urbana, Ill.

Stein, Irving. B.S. (Mass Inst. Tech.) Asso. Stat., War Dept., Washington, D. C. *611 Oglethorpe St.*

Stergion, Andrew P. M.S. (Mass Inst. Tech.) 1st Lt., USA, The Proving Center, Aberdeen Proving Gd., Md.

Sternhell, Arthur I. B.A. (New York) Staff Asst., Metropolitan Life Ins. Co., *1938 E. Tremont Ave., Parkchester, N. Y.*

Thompson, Louis T. E. Ph.D. (Clark) Dir. Res. and Dev., Lukas-Harold Corp., Indianapolis, Ind. *340 East Maple Rd.*

Tyler, Asst. Prof. George W. M.A. (Duke) Virginia Polytechnic Inst., Blacksburg, Va.

Working, Holbrook S. Ph.D. (Wisconsin) Chief Stat. Consultant, War Production Board, Washington, D. C. *Food Res. Inst., Stanford Univ., Calif.*

The following persons have been elected to Junior membership in the Institute:

Blumenthal, Lydia. Hunter College, New York, N. Y. *1001 Lincoln Pl., Brooklyn, N. Y.*

Gunlogson, Lee. Univ. of Minnesota, Minneapolis, Minn. *1906 Third Ave.*

Heacock, Richard R. Oregon State Coll., Corvallis, Ore. *P. O. Box 207, Seaside, Ore.*

Locatelli, Humbert J. Columbia Univ., New York, N. Y. *44 Seaman Ave.*

Mathisen, Harold C., Jr. Princeton Univ., Princeton, N. J. *4 Middle Dod Hall.*

Murphy, Ray Bradford. Princeton Univ., Princeton, N. J. *28 Godfrey Rd., Upper Montclair, N. J.*

Peters, Edward J., Jr. Georgetown Univ., Washington, D. C. *126 St. James Pl., Atlantic City, N. J.*

Smith, Joan T. Univ. of Minnesota, St. Paul, Minn. *673 East Nebraska Ave.*

SPECIAL COURSES IN STATISTICAL QUALITY CONTROL

The application of statistics to quality control is now being furthered in a program in which the War Production Board and the U. S. Office of Education are cooperating to assist statisticians in various industrial areas to provide suitable courses of instruction sponsored by their own institutions.

The general plan of the program has been influenced by two conclusions drawn from the experience gained in ESMWT courses carried on by Stanford University during 1942–43.[1] These conclusions were: (1) that a short full-time course in statistical quality control tends to be peculiarly effective; and (2) that it is vital to have the initial courses followed by meetings in which the course members gather to report on applications they have made and to receive encouragement and any needed assistance.

The giving of short full-time courses presents a problem of assembling a suitable staff, since four instructors will ordinarily be needed. If this problem were solved by arranging for a single staff to tour all the principal industrial regions giving courses in quality control, the local leadership necessary for establishing widespread use of statistical methods of quality control in industry would not be developed. The program adopted seems to offer an effective solution of these problems.

Under the program now in effect, the War Production Board, through its Office of Production Research and Development supplies an experienced person to assist with the arrangement of courses and to participate in the instruction.[2] Two of the instructors in each course will ordinarily be provided by a local educational institution, which will also promote the course and make necessary local arrangements through its institutional representative of the Engineering Science and Management War Training program. It is not considered necessary that the instructors provided by the institution have previous experience with statistical quality control provided they are sufficiently competent in the theory of sampling, but it is desirable that at least one of them have practical experience with quality control. It may often happen that one of the instructors can be a quality control man from a local industrial establishment. The representative of the WPB will assist with arrangements for bringing in one (or, where needed, two) additional outside instructors.

The sponsoring institution costs for the courses, which do not include the salary and expenses of the representative of the WPB, may be provided through the ESMWT program. The follow-up work with men who have taken the initial courses may be arranged also as part of the ESMWT program of the

---

[1] A description of these courses offered by Stanford University appeared in the *Annals of Mathematical Statictics*, March 1943, p. 96.

[2] At present Professor Holbrook Working is serving in this capacity.

educational institution sponsoring the original course. The follow-up work should be handled by a local instructor who participated in the original course.

The two basic courses and the one follow-up course that have already been given by Stanford University were conducted under essentially the plan outlined above, except that they did not have the benefit of assistance from the WPB. Three courses have thus far (May 25) been arranged under the new plan: one sponsored by Rhode Island State College, to be held during May 27 to June 2 at Newport, and two sponsored by Stanford University, to be held respectively in Los Angeles, June 13 to 20, and in San Francisco, June 22 to 29. Preliminary steps have been taken toward the arrangement of several additional courses.

REPORT OF THE NEW YORK MEETING OF THE INSTITUTE

A joint meeting between the Institute and the American Society of Mechanical Engineers was held on Saturday, May 29, 1943 at the Engineering Societies Building, 29 West 39th Street, New York City. Of the ninety-five individuals attending the meeting, the following fifty-seven members of the Institute were present:

Theodore W. Anderson, K. J. Arnold, Robert E. Bechhofer, B. M. Bennett, C. I. Bliss, Mary E. Boozer, P. Boschan, A. H. Bowker, Burton H. Camp, A. C. Cohen Jr., H. F. Dodge, C. Eisenhart, Mary L. Elveback, W. C. Flaherty, H. Goode, John I. Griffin, Charles C. Grove, Frank E. Grubbs, E. J. Gumbel, Harold Hotelling, J. M. Juran, B. F. Kimball, Lila Knudsen, Howard Levene, E. Vernon Lewis, Simon Lopata, Frank W. Lynch, Henry Mann, E. C. Molina, N. Morrison, Philip J. McCarthy, Luis F. Nanni, Franklin S. Nelson, M. L. Norden, P. S. Olmstead, R. F. Passano, Edward Paulson, G. A. D. Preinreich, A. C. Rosander, Arthur Sard, Henry Scheffé, Bernice Scherl, Edward M. Schrock, L. W. Shaw, William B. Simpson, S. G. Small, Arthur Stein, Andrew P. Stergion, M. Stevens, David F. Votaw Jr., A. Wald, Helen M. Walker, W. A. Wallis, S. S. Wilks, J. Wolfowitz, L. C. Young.

The general topic of the meeting was *Industrial Applications of Statistics.* At the morning session the following papers were presented, with Professor Harold Hotelling presiding:

1. *On the Theory of Runs with some Application to Quality Control.*
   J. Wolfowitz.
2. *On the Presentation of Data as Evidence.*
   Churchill Eisenhart.

At the afternoon session, the following papers were presented with Mr. E. C. Molina as Chairman:

1. *A Sampling Inspection Plan for Continuous Production.*
   H. F. Dodge.
2. *Tolerances and Product Acceptability.*
   L. C. Young.

A meeting of the Board of Directors was held after the afternoon session.

EDWIN G. OLDS
Secretary

Photo-Lithoprint Reproduction
EDWARDS BROTHERS, INC.
Lithoprinters
ANN ARBOR, MICHIGAN
1947